CHARACTERISATION OF WHOLE WHITE MAIZE KERNELS USING SPECTRAL IMAGING

by

Kate Sendin

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science
in the Faculty of AgriScience at Stellenbosch University

Supervisor: Prof. Marena Manley

Co-supervisor: Dr Paul J. Williams

March 2017

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Kate Sendin

March 2017

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisors Prof. Manley and Dr. Williams for the continuous support of my research. Their patience, motivation, and vast knowledge were greatly appreciated whilst conducting the research and throughout the writing of the thesis. Under your guidance I have grown immeasurably as a researcher.

I would like to extend my most sincere thanks to the following people and institutions for their contribution to the successful completion of this study:

Prof. Geladi (Swedish University of Agricultural Sciences) and Prof. Linderholm (Umeå Univeristy) for the use of their Videometer and Sisuchema instruments, and for their expert advice while introducing me to the practical side of hyperspectral imaging;

Wiana Louw and the graders at the South African Grain Laboratory for going out of their way to gather samples on my behalf;

All staff and post-graduate students in the Food Science Department for a warm and friendly working environment. Special thanks to the head of our department, Prof. Sigge;

Megan van der Merwe for kindly translating my thesis abstract to Afrikaans – jy's 'n engel; and

The National Research Fund (NRF) (Thutuka Bursary, 2015, and Scarce Skills Bursary, 2016) are hereby acknowledged for financial support (any opinion, findings and conclusions or recommendations expressed in this material are those of the author and therefore the NRF does not accept any liability in regard thereto).

Last but not least, I would like to thank my parents Billy and Heidi, my brother William, and Charles Peek for their continued support and encouragement throughout writing this thesis.

ABSTRACT

Maize (*Zea mays L.)* is the most important cereal crop grown in South Africa. It is produced widely across the country under diverse environments, and thus a variety of defects tend to occur. Grading is an important quality and safety control step where these defective materials are identified and quantified. This study considered the most important defective material classes, namely 6 types of defective white maize kernels, 5 types of foreign matter, other colour kernels (yellow maize) and pinked white maize kernels. Current maize grading is manual and tedious, and modern analytical methods could improve this process. This study aimed to investigate the viability of using spectral imaging with multivariate data analysis for maize grading by separating sound maize from the 13 defective materials classes.

NIR hyperspectral imaging with pixel-wise and object-wise data analysis were used for two-way discrimination of the sound and defective material classes. The average spectra indicated prominent bands at 1219 and 1476 (related to starch), 1941 (related to moisture), and 2117 nm (related to protein). The loadings of principal component (PC) 1 exhibited similar bands. The object-wise approach performed superiorly to the pixel-wise approach across all 13 analyses. Little separation was observed in the principal component analysis (PCA) score plots in the pixel-wise results due to a large similarity between classes. The object-wise approach utilised the average spectrum for each maize kernel, and the overlap was reduced. Partial least squares discriminant analysis (PLS-DA) models were calculated and used to classify an independent validation set of 30 sound kernels and 30 defective materials. The pixel-wise analyses achieved classification accuracies ranging 75-99%. This approach was not able to accurately distinguish closely related classes. The object-wise analyses performed well, with 8 of the 13 achieving 100% classification accuracy, and the remaining 5 classes incurring only one error per analysis of 60 kernels.

Multispectral imaging followed to compare the two imaging techniques. Pixel-wise PCA was applied to pre-process the spectral imaging data, followed by object-wise two-way PLS-DA modelling using 17 sound kernels and 18 defective material objects. The PCA loadings revealed that colour played a role in separating the classes, with a wide band appearing across 505, 525, 570 and 590 nm. Classification accuracies of 83-100% were achieved, and were generally slightly lower than the results obtained for all classes using the NIR hyperspectral imaging instrument.

Spectral imaging was shown to be capable of separating white maize from 13 commonly occurring defective materials. NIR hyperspectral imaging performed superiorly to multispectral imaging, and the use of an object-wise data analysis approach further improved the accuracy of the separations. These techniques have the potential to offer the maize industry a rapid, accurate and objective alternative grading method.

UITTREKSEL

Mielies (*Zea mays L.*) is die belangrikste graangewas wat tans in Suid Afrika geproduseer word. Dit word landwyd geproduseer, en deurdat dit in soveel diverse omgewings groei, word verskeie defekte gereeld opgespoor. Gradering is ‘n baie belangrike kwaliteit- en veiligheidsmaatreël waardeur hierdie foutiewe materiaal uitgesonder en dan gekwantifiseer word. Hierdie studie oorweeg die mees prominente foutiewe materiaal klasse, naamlik: ses tipes foutiewe wit mieliepitte, vyf tipes vreemde materiaal, anderskleurige mieliepitte (geel mielies) en verpienkte wit mieliepitte. Huidige mieiliegradering is ‘n duur en tydsame proses, en moderne analitiese metodes kan hierdie proses verbeter. Hierdie studie stel ondersoek in om te bepaal asof die gebruik van spektrale beelding met meerveranderlike data ontleding vir mieliegradering lewensvatbaar is, deur gesonde mielies van die 13 foutiewe materiaal klasse te skei.

Naby infrarooi (NIR) hiperspektrale beelding met pixel- en voorwerp-wyse data analiese is gebruik vir ‘n tweerigting diskriminasie van die gesonde en foutiewe materiaal klasse. Die gemiddelde spektra het prominente bande aangedui by 1219 en 1476 (stysel-verwant), 1941 (proteien-verwant) en 2117 nm (vog-verwant). Die lading-stip van hoofkomponent (HK) 1 het soortgelyke bande gewys. Die voorwerp-wyse benadering het regoor al 13 analises beter as die pixel-wyse benadering presteer. As gevolg van ‘n groot ooreenkomste tussen die verskillende klasse, was min skeiding geobserveer in die hoofkomponent analise (HKA) telling-beelde in die pixel-wyse resultate. Die voorwerp-wyse benadering het van die gemiddelde spektrum van elke mieliepit gebruik gemaak, en die oorvleuling was so verminder. Parsiële kleinste waarde diskriminantanalise (PKW-DA) modelle was bereken om 30 gesonde- en 30 foutiewe materiale te klassifiseer. Die pixel-wyse analises het klassifikasie akkuraatheid tussen 75-99% bereik. Hierdie benadering kon nie akkuraat tussen die verwante klasse onderskei nie. Die voorwerp-wyse analise het goed presteer, waar 8 van die 13, 100% klasifikasie akkuraatheid bereik het, en die oorblywende 5 klasse net een fout per analise van 60 pitte aangegaan het.

Multispektrale beelding het gevolg om die twee beeldingstegnieke te vergelyk. Pixel-wyse HKA was bereken om skoonmaak van die beeld te akkomodeer, en was vervolg deur voorwerp-wyse tweerigting PKW-DA modellering wat van 17 gesonde pitte en 18 voorwerpe van foutiewe materiaal gebruik gemaak het. Die HKA lading-stippe het onthul dat kleur ‘n massiewe rol in die skeiding van die klasse gespeel het, met ‘n wye band wat oor 505, 525, 570 en 590 nm verskyn het. Klassifikasie akkuraatheid van 83-100% was bereik, en was oor die algemeen iewat laer as die resultate wat in alle klasse bereik is tydens die gebruik van die NIR hiperspektrale beelding instrument.

Dit was daardeur gewys dat spektrale beelding bekwaam was om wit mielies van 13 bekende foutiewe materiale te skei. NIR hiperspektrale beelding het beter gedoen as multispektrale beelding, en die gebruik van ‘n voorwerp-wyse data analise benadering het verder die akkuraatheid van die skeidings verbeter. Hierdie tegnieke het die potensiaal om vir die mielie industrie ‘n vinnige, akkurate en objektiewe alternatiewe graderings metode aan te bied.

TABLE OF CONTENTS

DECLARATION i
ACKNOWLEDGMENTS ii
ABSTRACT iii
UITREKSEL iv
LIST OF FIGURES vii
LIST OF TABLES xii
LIST OF ABBREVIATIONS USED xiii
CHAPTER 1: INTRODUCTION 1
References 4
CHAPTER 2: LITERATURE REVIEW 8
Introduction 9
Maize Grading in South Africa 11
Brief History of Hyperspectral Imaging 14
Principles and Theoretical Background of Hyperspectral Imaging 14
Fundamentals of Hyperspectral Imaging 14
Components of Hyperspectral Imaging Systems 16
Acquisition of Hyperspectral Images 17
Analysis of Hyperspectral Images 18
Applications of Hyperspectral Imaging in Cereal Evaluation 21
Quality 21
Safety 30
Conclusion 34
References 34
CHAPTER 3: CHARACTERISATION OF WHITE MAIZE KERNELS USING NIR HYPER-SPECTRAL IMAGING 42
Introduction 43
Materials and Methods 44
Samples 44
NIR Hyperspectral System 45
Image Acquisition 45
Hyperspectral Image Analysis 46
Results and Discussion 49

Spectral Analysis 49
Multivariate Data Analysis 51
Conclusion 68
References 68
CHAPTER 4: CHARACTERISATION OF WHITE MAIZE KERNELS USING MULTI-SPECTRAL IMAGING 72
Introduction 73
Materials and Methods 74
Samples 74
Multispectral System 74
Image Acquisition 75
Multispectral Image Analysis 75
Results and Discussion 78
Multivariate Data Analysis 78
Conclusion 90
References 91
CHAPTER 5: GENERAL DISCUSSION AND CONCLUSIONS 94
References 96
ADDENDUM 1: SOUTH AFRICAN MAIZE GRADING REGULATIONS 98

LIST OF FIGURES

Figure 2.1 A NIR hyperspectral imaging hypercube comprises of one wavelength (z) and two spatial (x and y) dimensions. One can view the data as a spectrum of a pixel in the sample, or as an image plane of the entire sample at a chosen wavelength (e.g. at 1000 nm). 15

Figure 2.2 Schematic of a pushbroom NIR hyperspectral imaging system, illustrating the line-by-line data acquisition of the samples on the linear translation stage. 16

Figure 2.3 (a) Scores images of PC2 and PC5 for whole yellow maize kernels enabling visualisation of similarity in chemical composition (similar colours indicate similar chemical composition, in this case similar endosperm texture). (b) Scores plot of PC2 vs. PC5 with three clusters. (c) Classification plot based on clusters identified in the PC scores plot. (d) Classification image after projection of the classes identified in the scores plot onto the scores image [Copyright 2014 Royal Society of Chemistry. Reproduced with permission from Manley (2014)]. 19

Figure 2.4 (a) PC5 score image for the barley cultivar Erica image dataset where dark blue indicates viable kernels and green non-viable kernels. (b) The corresponding PC5 vs. PC1 score plot showing density and clustering of pixels. Positive PC5 scores values (green box) were identified and assigned to the viable class and negative PC5 scores values (blue box) to the non-viable class. Brushing between score plot and score image was used. (c) Classification image showing viable and non-viable regions. [Copyright 2011 Springer. Reproduced with permission from McGoverin et al. (2011)]. 29

Figure 2.5 (a) PCA score plot of PC4 vs. PC5 (0.49% and 0.34%) showing three clusters and (b) corresponding score image of PC4 showing decrease in score values from left to right, where warm colours (yellow to red) correspond with positive score values and cold colours (blue) correspond to negative score values; (c) classification plot of PC4 vs. PC5 with classes and degree of infection in the direction of PC4 and (d) classes projected onto score image showing the control (green), T0&T1 (black) and the remaining time intervals up to 90 h post inoculation (red). [Copyright 2012 Elsevier. Reproduced with permission from Williams et al. (2012)]. 31

Figure 3.1 Digital image of (a) sound white maize and the 13 undesirable materials: (b) *Fusarium* damage; (c) *Diplodia* damage; (d) pinked maize; (e) water damage; (f) rodent damage; (g) heat damage; (h) plant material; (i) screenings; (j) wheat; (k) sorghum; (l) soy; (m) sunflower; and (n) yellow maize. 45

Figure 3.2 Summary of image analysis process, illustrated using sound class (top) vs. yellow class (bottom) with the pixel-wise approach (object-wise analysis followed the same method). The left half of each section (a–f) is germ-up, and the right half is germ-down. Each half contains the same kernels in the same positions, to allow for inspection of both sides. (a) Raw mosaic image (at 1426 nm) of calibration samples arranged in known order of classes; (b) PCA score image of calibration samples after pre-processing; (c) Samples were assigned classes (sound as green and defect as blue) and PLS-DA model was calibrated; (d) Raw mosaic image (at 1426 nm) of independent validation samples arranged in known order of classes; (e) PCA score image of validation samples after pre-processing; and (f) PLS-DA model applied to generate classification image. 46

Figure 3.3 Method used for totalling the object-wise classification results, illustrated using yellow maize class vs. sound class. (a) Unaltered object-wise classification image; (b) Assignment of overall classification; and (c) Overall object-wise classification image. 48

Figure 3.4 Unprocessed average pseudo-absorbance spectra for sound white maize; white maize defects (heat damage, rodent damage, *Diplodia* damage, water damage, screenings and *Fusarium* damage); and pinked white maize. 50

Figure 3.5 Unprocessed average pseudo-absorbance spectra for sound white maize; yellow maize; and foreign matter (sunflower seeds, sorghum, wheat, plant material and soy). 51

Figure 3.6 Pixel-wise PCA of heat damage class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (79% SS) vs. PC2 (8% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 51

Figure 3.7 Pixel-wise PCA of wheat class vs. sound class. Slight separation of classes was observed. Scores given as (a) PCA score plot of PC1 (81% SS) vs. PC2 (7% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 52

Figure 3.8 Pixel-wise PCA of yellow maize class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (82% SS) vs. PC2 (7% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 52

Figure 3.9 Pixel-wise PCA of screenings class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (79% SS) vs. PC2 (6% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 52

Figure 3.10 Pixel-wise PCA of *Diplodia* damage class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (82% SS) vs. PC2 (6% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 53

Figure 3.11 Pixel-wise PCA of pinked class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (81% SS) vs. PC2 (8% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2. 53

Figure 3.12 Object-wise PCA of heat damage class vs. sound class. Minimal separation of classes was observed. (a) PCA score plot of PC1 (78% SS) vs. PC2 (12% SS). (b) Classes of sound (green) and heat damaged (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 54

Figure 3.13 Object-wise PCA analysis of wheat class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (74% SS) vs. PC2(15% SS). (b) Classes of sound (green) and wheat (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 55

Figure 3.14 Object-wise PCA analysis of yellow maize class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (80% SS) vs. PC2 (13% SS). (b) Classes of sound (green) and yellow maize (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 55

Figure 3.15 Object-wise PCA analysis of screenings class vs. sound class. Good separation of classes was observed. (a) PCA score plot of PC1 (70% SS) vs. PC2 (15% SS). (b) Classes of sound (green) and screenings (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 54

Figure 3.16 Object-wise PCA analysis of *Diplodia* damage class vs. sound class. Slight separation of classes observed. (a) PCA score plot of PC1 (86% SS) vs. PC2 (8% SS). (b) Classes of sound (green) and *Diplodia* damage (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 56

Figure 3.17 Object-wise PCA analysis of pinked class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (85% SS) vs. PC2 (9% SS). (b) Classes of sound (green) and pinked (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2. 56

Figure 3.18 Pixel-wise PLS-DA classification resulted in well classified heat damage class vs. sound class (90.72% classification accuracy). Fair separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (78% SS) vs. 2 (6% SS); (b) Classification image. 57

Figure 3.19 Pixel-wise PLS-DA classification resulted in extremely poorly classified wheat class vs. sound class (90.87% classification accuracy; 2.38% sensitivity). Minimal separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (72% SS) vs. 2 (23% SS); (b) Classification image. 59

Figure 3.20 Pixel-wise PLS-DA classification resulted in poorly classified yellow maize class vs. sound class (75.32% classification accuracy). Minimal separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (81% SS) vs. 2 (7% SS); (b) Classification image. 59

Figure 3.21 Pixel-wise PLS-DA classification resulted in well classified screenings class vs. sound class (91.58% classification accuracy). Fair separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (64% SS) vs. 2 (14% SS); (b) Classification image. 60

Figure 3.22 Pixel-wise PLS-DA classification resulted in well classified *Diplodia* damage class vs. sound class (85.22% classification accuracy). Slight separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (81% SS) vs. 2 (5% SS); (b) Classification image. 60

Figure 3.23 Pixel-wise PLS-DA classification resulted in extremely poorly classified pinked class vs. sound class (62.94% classification accuracy). Slight separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (74% SS) vs. 2 (12% SS); (b) Classification image. 61

Figure 3.24 Object-wise PLS-DA classification resulted in perfectly classified heat damage class vs. sound class (100% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (75% SS) vs. 2 (8% SS); (b) Unprocessed classification image; and (c) Overall classification image. 63

Figure 3.25 Object-wise PLS-DA classification resulted in perfectly classified wheat class vs. sound class (100% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (71% SS) vs. 2 (14% SS); (b) Unprocessed classification image; and (c) Overall classification image. 64

Figure 3.26 Object-wise PLS-DA classification resulted in well classified yellow maize class vs. sound class (98.33% classification accuracy). (a) PLS-DA score plot PLS factor 1 (79% SS) vs. 2 (12% SS); (b) Unprocessed classification image; and (c) Overall classification image. 64

Figure 3.27 Object-wise PLS-DA classification resulted in perfectly classified screenings class vs. sound class (100% classification accuracy). (a) PLS-DA score plot PLS factor 1 (47% SS) vs. 2 (25% SS); (b) Unprocessed classification image; and (c) Overall classification image. 64

Figure 3.28 Object-wise PLS-DA classification resulted in perfectly classified *Diplodia* damage class vs. sound class (100% classification accuracy). (a) PLS-DA score plot PLS factor 1 (83% SS) vs. 2 (9% SS); (b) Unprocessed classification image; and (c) Overall classification image. 65

Figure 3.29 Object-wise PLS-DA classification resulted in well classified pinked class vs. sound class (98.15% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (73% SS) vs. 2 (12% SS); (b) Unprocessed classification image; and (c) Overall classification image. 65

Figure 4.1 Summary of image analysis, illustrated using sound class (first 17 kernels) vs. yellow class (second 18 kernels). The top half of each section (a–f) is germ-up, and the bottom half is germ-down. Each half contains the same kernels in the same positions, to allow for inspection of

both sides. (a) Digital image of calibration samples arranged in known order of classes; (b) PCA score image of calibration samples after pre-processing; (c) Samples were assigned classes (sound as green and defect as blue) and PLS-DA model was calibrated; (d) Digital image of independent validation samples arranged in known order of classes; (e) PCA score image of validation samples after pre-processing; and (f) PLS-DA model applied to generate classification image. 76

Figure 4.2 Method used for totalling results, illustrated using *Diplodia* class vs. sound class. (a) Unprocessed classification image; (b) Totalling procedure for determining overall classification; and (c) Overall classification image. 77

Figure 4.3 Well separated heat damage class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses. 78

Figure 4.4 Well separated wheat class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses. 78

Figure 4.5 Well separated yellow maize class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 do not represent the general loadings line trend observed over most of the 13 PCA analyses. 79

Figure 4.6 Moderately separated screenings class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the inverse of the general loadings line trend observed over most of the 13 PCA analyses. 79

Figure 4.7 Not separated *Diplodia* damage class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses. 80

Figure 4.8 Not separated pinked class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 do not represent the general loadings line trend observed over most of the 13 PCA analyses. 80

Figure 4.9 Perfectly classified heat damage class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 85

Figure 4.10 Perfectly classified wheat class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 85

Figure 4.11 Perfectly classified yellow maize class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 86

Figure 4.12 Perfectly classified screenings class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 86

Figure 4.13 Poorly classified *Diplodia* damage class vs. sound class (82.86% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 87

Figure 4.14 Moderately well classified pinked class vs. sound class (91.43% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image. 87

LIST OF TABLES

Table 2.1 Standards for grades of white maize (per category of undesirable material and maximum allowed weight) according to South African grading regulations. 12

Table 2.2 Applications of hyperspectral imaging and chemometrics for quality and safety evaluation of cereals and cereal products. 22

Table 3.1 Method used to determine the overall classification of a kernel based on the classification image generated from the germ-up and -down image data. 48

Table 3.2 Results of pixel-wise PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes. 62

Table 3.3 Results of object-wise PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes. 66

Table 4.1 The wavebands utilised by the VideometerLab2 instrument and the associated region in the electromagnetic spectrum. 75

Table 4.2 Results of PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes. 88

LIST OF ABBREVIATIONS USED

3D	three dimensional
ANN	artificial neural networks
AOTF	acousto-optic tunable filter
AVRIS	airborne visible/infrared imaging spectrophotometer
BPNN	back propagation neural network
CCD	charged coupled device
CMOS	complementary metal oxide semiconductor
FDA	factorial discriminant analysis
HgCdTe	mercury-cadmium-telluride
InGaAs	indium-gallium-arsenide
LCTF	liquid crystal tunable filter
LDA	linear discriminant analysis
LDC	linear discriminant classifier
LED	light emitting diode
MIR	mid infrared
MLR	multiple linear regression
MNF	minimum noise fraction
MSC	multiplicative scatter correction
MSEP	mean square errors of prediction
NASA/JPL	National Aeronautics and Space Administration Jet Propulsion Laboratory
NIR	near infrared
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PDA	penalised discriminant analysis
PLS	partial least squares
PLS-DA	partial least squares discriminant analysis
PLS-FDA	partial least squares factorial discriminant analysis
PLSR	partial least squares regression
Q^2	cross-validated coefficient of determination
QDA	quadratic discriminant analysis
R^2	coefficient of determination
RMSEP	root mean square error of prediction
SAGL	South African Grain Laboratory
SAM	spectral angle mapping

SECV	standard error of cross-validation
SNV	standard normal variate
SS	sum of squares
SVM	support vectors machine
SWIR	short wave infrared
UV	ultra-violet
WM1	white maize grade 1
WM2	white maize grade 2
WM3	white maize grade 3

CHAPTER 1
INTRODUCTION

Maize (*Zea mays L.*) is a widely consumed and grown cereal crop. It can be found in many forms such as minimally processed staple foods (porridges and breads), processed foods (tortillas, chips and breakfast cereals), and even in alcohol production (Fox and Manley, 2009). Generally, developing regions, such as in Africa, South America and Asia, consume maize directly as staple diet (Fox and Manley, 2009). In developed countries, maize is more often consumed as second-cycle produce, in the form of meat, eggs and dairy (Scanes, 2010). Maize, particularly white maize, is a vital food source for African people, including the majority of the South African population (Gouse *et al.*, 2006). In South Africa, white maize is generally consumed as a minimally processed unfermented porridge, which is commonly the base of all meals. Although maize lacks one essential amino acid, lysine, it is a good source of protein and carbohydrates, and can constitute a large proportion of a healthy diet (Eckhoff and Paulsen, 2012). The popularity of maize as a food crop is a combination of its ease of cultivation, large yields, storage ability, and high content of easily metabolised starch (Eckhoff and Paulsen, 2012). Furthermore, the low cost, reliable supply, high digestibility, and close proximity of maize growing and livestock rearing regions contribute to maize products generally being the most cost-effective animal feed available.

During maize classification several physical properties of a sample are evaluated to give an indication of a sample's chemical composition, functionality and appropriate end-uses (Serna-Saldivar, 2010). First, the class is determined by maize kernel colour, white or yellow. Each class is then graded in accordance with a class-specific grading regulation. During grading, the main aim is to determine the grain condition and health. Regardless of the specific product being produced, certain grading quality and safety characteristics are important for all human food uses (Eckhoff and Paulsen, 2012). The maize must not exhibit visual insect, rodent or fungal deterioration, as the palatability, sanitation and shelf life of the resulting products would be decreased. Furthermore, mycotoxins originating from fungal contamination could cause sickness or death if consumed (Yoshizawa *et al.*, 1994; Sydenham *et al.*, 1990). The kernels should be intact, and the pericarp should not exhibit obvious damage. Broken and damaged maize will have different water absorbing properties, leading to less uniform cooking characteristics and final product consistency. The grading of maize facilitates fair marketing between buyers and sellers by providing an indication of its intrinsic value. Subsequent to grading, further classification provides more insight into industrial use or functionality. Factors such as test weight (hectolitre mass), hardness, or chemical composition are not considered during grading, but may be agreed upon in an additional trade agreement.

While many manual methods are utilised in the cereal industry, especially in grading, much research has focused on alternative analytical methods that aim to avoid human error and subjectivity. Considerable work has focused on the development of non-invasive and non-destructive techniques. These include reflectance and transmittance spectroscopy (including ultraviolet, visible, near infrared (NIR) and mid infrared (MIR)), acoustic response (Pearson *et al.*, 2007; Stasiak *et al.*, 2007), and X-ray computed tomography (Neethirajan *et al.*, 2008; Zhu *et al.*, 2012). Spectroscopic techniques, most commonly utilising the NIR region, have shown to be promising for the non-destructive and non-invasive inspection of cereals (Fox and Manley, 2014; Sendin *et al.*, 2016). NIR spectroscopy is based on the bond vibrations within organic molecules consisting mainly of C–H, C=O, O–H and N–H, which occur due to excitement by light with wavelengths of 780 to 2500 nm (Manley, 2014). NIR spectroscopy is widely implemented in the cereal industry, and has been found useful in qualitative and quantitative measuring of major components, including dry matter, moisture, protein and starch, as well as minor components, such as gluten (Huang *et al.*, 2008). Applications include inspection pre- and during harvest (Engel *et al.*, 1997; Maertens *et al.*, 2004; Montes *et al.*, 2006), during grain processing (Kawamura *et al.*, 2003), and in cereal end-product quality (De Temmerman *et al.*, 2007; Kays *et al.*, 1996; Weeranantanaphan and Downey, 2010). Although studies have demonstrated that conventional spectroscopy is capable of evaluating a range of maize properties, it has some limitations. Foods and biomaterials are, in general, heterogeneous and it is often a challenge to obtain spectral data that are representative for the chemical composition and distribution. No information is gathered in the spatial dimension, and thus measurements may only cover a small area of the kernel. As maize kernels are heterogeneous, measurements of the entire kernel are necessary. In contrast, spectral imaging offers localisation, allowing entire samples to be analysed rapidly and accurately.

Spectral imaging, known also as chemical imaging, integrates conventional imaging and spectroscopy to attain both spatial and spectral information from the sample (Gowen *et al.*, 2007). A spectral image, commonly referred to as a three-dimensional (3D) hypercube, comprises numerous adjacent wavelengths for each spatial point of the object. Thus, each pixel in the image contains the spectrum at that exact point (Burger and Geladi, 2006). Multispectral imaging and hyperspectral imaging are two spectral imaging techniques based on the same principles. The difference between the two is simply the complexity of the spectral data collected. Hyperspectral imaging systems acquire spectra in hundreds of narrow bands that are separated by regular intervals (e.g. one band every 5 nm). The resulting spectra are smooth and closely resemble the spectra of conventional techniques. Multispectral imaging systems often utilise a handful of wavebands, *ca*. 20, which are irregularly spaced across the chosen range, and thus a jagged spectrum is achieved. Multispectral imaging systems are simpler, and thus offer faster acquisition times and lower cost of the camera. Hyperspectral imaging is more suited to detailed research and laboratory applications, while

multispectral imaging offers a more practical solution to industry requirements. Both techniques will be explored in this study.

To establish if spectral imaging has the potential for replacing manual maize grading, one must understand the current demands of the maize industry in South Africa. Refer to Addendum 1 for the relevant maize grading legislation. In manual grading, a sample (*ca.* 150 g) of a maize consignment is visually inspected to identify any content other than sound white maize, such as defective or foreign material (Department of Agriculture, 2009). The method involves spreading the kernels onto a flat surface and slowly inspecting each kernel. Any undesirable materials are weighed and referenced to maximum levels stipulated for each material's category. There are three white maize classes, WM1 (best), WM2 and WM3. The grade is allocated solely on the contents of undesirable materials within a sample. Unfortunately, evaluating physical features by human inspection requires significant training and is very tedious, even on small samples. To emulate these current practices and demonstrate that sound (healthy) maize could be separated from all of the undesirable materials using an objective method such as spectral imaging, two-way comparisons must be made. The two classes, sound white maize and a single defect, must be separated through a series of analyses. Furthermore, the grader inspects both sides (germ-up and germ-down) of every kernel to detect defects during manual grading, as the extent and presence of a defect on individual kernels is extremely variable and irregular. Thus, identical germ-up and germ-down image pairs must be collected for assessing each side of all kernels.

The analytical capability of the near infrared (NIR) spectroscopy component of spectral imaging relies on absorption by wavebands that often overlap, making interpretation of spectra difficult. In addition, the sheer size of the hypercube further complicates data analysis and interpretation. By using multivariate data analysis, the composition of the object may be charaterised by relating the varying spectra to their spatial distributions. Accurate calibration equations can be calculated for various cereal quality and safety attributes. Before being subjected to these chemometric analyses, data must first be mathematically pre-processed. Analytical information is extracted from spectra, and non-chemical biases, such as scattering effects, are removed (Del Fiore *et al.*, 2010). Principal component analysis (PCA) is arguably one of the most useful and widely used exploratory data analysis techniques (Amigo *et al.*, 2013; Cowe and McNicol, 1985). It is used to generate score and loading plots that may be visually assessed. Once data has been explored using PCA, a popular supervised classification technique, partial least squares discrimination analysis (PLS-DA), may be applied. This combination of PCA and PLS-DA has been used effectively to identify, differentiate or evaluate a variety of cereal properties (Cogdill *et al.*, 2004; Del Fiore *et al.*, 2010; Manley *et al.*, 2009; McGoverin *et al.*, 2011; Vermeulen *et al.*, 2012; Wallays *et al.*, 2009; Williams *et al.*, 2009).

Multivariate data analysis traditionally folllows a pixel-wise approach (Williams *et al.*, 2012; McGoverin and Manley, 2012). However, an object-wise approach has emerged in recent years

(Kucheryavskiy, 2013; Williams and Kucheryavskiy, 2016). Pixel-wise analysis simply utilises each pixel's spectrum as a sample for classification. However, during object-wise image analysis an average spectrum from all pixels that fall within a single object is calculated (i.e. each maize kernel). The average spectrum is used for classification instead of original spectra of individual pixels. When different classes of samples are closely related, such as in maize kernels, the majority of pixels exhibit similar spectral responses. A large number of misclassified pixels often occur when using a pixel-wise approach. This can be avoided by classifying objects to significantly decrease the amount of data and calculate more stable classification models. As few studies have compared the two approaches, the best approach for separating closely related healthy and defective maize, as well as other cereal commodities and foreign matter, must be identified.

This study investigated the potential for maize grading by spectral imaging. Two imaging systems, NIR hyperspectral imaging (1118–2425 nm) and multispectral imaging (375–970 nm), as well as two multivariate data analysis approaches, pixel-wise and object-wise, were compared. Specific objectives were to establish capability of:

- NIR hyperspectral imaging to separate sound white maize kernels from common grading defects; and
- multispectral imaging to achieve comparable results for separating sound white maize kernels from common grading defects.

REFERENCES

Amigo, J. M., Martí, I. & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality. In: *Data Handling in Science and Technology*. Pp. 343-370. Amsterdam, The Netherlands: Elsevier Science.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst,* **131,** 1152-1160.

Cogdill, R. P., Hurburgh, C. R., Rippke, G. R., Bajic, S. J., Jones, R. W., McClelland, J. F., Jensen, T. C. & Liu, J. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE,* **47,** 311.

Cowe, I. A. & McNicol, J. W. (1985). The use of principal components in the analysis of near-infrared spectra. *Applied Spectroscopy,* **39,** 257-266.

De Temmerman, J., Saeys, W., Nicolaï, B. & Ramon, H. (2007). Near infrared reflectance spectroscopy as a tool for the in-line determination of the moisture concentration in extruded semolina pasta. *Biosystems Engineering,* **97,** 313-321.

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology,* **144,** 64-71.

Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Eckhoff, S. R. & Paulsen, M. R. (2012). Maize. In: *Cereal Grain Quality* (edited by R. Henry & P. Kettlewell). Pp. 77-112. London, UK: Springer Science & Business Media.

Engel, R., Long, D. & Carlson, G. (1997). On-the-go grain protein sensing is near. *Better Crops with Plant Food,* **81,** 20-23.

Fox, G. & Manley, M. (2009). Hardness methods for testing maize kernels. *Journal of Agricultural and Food Chemistry*, **57**, 5647-5657.

Fox, G. & Manley, M. (2014). Applications of single kernel conventional and hyperspectral imaging near infrared spectroscopy in cereals. *Journal of the Science of Food and Agriculture*, **94**, 174-179.

Gouse, M., Pray, C., Schimmelpfennig, D. & Kirsten, J. (2006). Three seasons of subsistence insect-resistant maize in South Africa: have smallholders benefited? *AgBioForum*, **9**, 15-22.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G. & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Huang, H., Yu, H., Xu, H. & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering,* **87,** 303-313.

Kawamura, S., Natsuga, M., Takekura, K. & Itoh, K. (2003). Development of an automatic rice-quality inspection system. *Computers and Electronics in Agriculture,* **40,** 115-126.

Kays, S. E., Windham, W. R. & Barton, F. E. (1996). Prediction of total dietary fiber in cereal products using near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry,* **44,** 2266-2271.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126-135.

Maertens, K., Reyns, P. & De Baerdemaeker, J. (2004). On-line measurement of grain quality with NIR technology. *Transactions of the ASAE,* **47,** 1135.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, **43**, 8200-8214.

Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (Zea maize L.) kernels. *Journal of Agricultural and Food Chemistry,* **57,** 8761-8769.

McGoverin, C. & Manley, M. (2012). Classification of maize kernel hardness using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy,* **20,** 529.

McGoverin, C. M., Engelbrecht, P., Geladi, P. & Manley, M. (2011). Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. *Analytical and Bioanalytical Chemistry,* **401,** 2283-2289.

Montes, J., Utz, H., Schipprack, W., Kusterer, B., Muminovic, J., Paul, C. & Melchinger, A. (2006). Near‐infrared spectroscopy on combine harvesters to measure maize grain dry matter content and quality parameters. *Plant Breeding,* **125,** 591-595.

Neethirajan, S., Jayas, D. S., White, N. D. G. & Zhang, H. (2008). Investigation of 3D geometry of bulk wheat and pea pores using X-ray computed tomography images. *Computers and Electronics in Agriculture*, **63**, 104-111.

Pearson, T. C., Cetin, A. E., Tewfik, A. H. & Haff, R. P. (2007). Feasibility of impact-acoustic emissions for detection of damaged wheat kernels. *Digital Signal Processing*, **17**, 617-633.

Scanes, C. (2010). Animal Nutrition: Feeds and Feeding. In: *Fundamentals of Animal Science*. Pp. 261-286. New York, USA: Cengage Learning.

Sendin, K., Williams, P. J. & Manley, M. (2016). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical Reviews in Food Science and Nutrition*.

Serna-Saldivar, S. O. (2010). Physical Properties, Grading, and Speciality Grains. In: *Cereal grains: properties, processing, and nutritional attributes*. Pp. 43-80. Boca Raton, USA: CRC Press.

Stasiak, M., Molenda, M. & Horabik, J. (2007). Determination of modulus of elasticity of cereals and rapeseeds using acoustic method. *Journal of Food Engineering*, **82**, 51-57.

Sydenham, E. W., Thiel, P. G., Marasas, W. F., Shephard, G. S., Van Schalkwyk, D. J. & Koch, K. R. (1990). Natural occurrence of some Fusarium mycotoxins in corn from low and high esophageal cancer prevalence areas of the Transkei, Southern Africa. *Journal of Agricultural and Food Chemistry*, **38**, 1900-1903.

Vermeulen, P., Pierna, J. A. F., Van Egmond, H. P., Dardenne, P. & Baeten, V. (2012). Online detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging. *Food Additives & Contaminants: Part A,* **29,** 232-240.

Wallays, C., Missotten, B., De Baerdemaeker, J. & Saeys, W. (2009). Hyperspectral waveband selection for on-line measurement of grain cleanness. *Biosystems Engineering,* **104,** 1-7.

Weeranantanaphan, J. & Downey, G. (2010). Identity confirmation of a branded, fermented cereal product by UV spectroscopy: A feasibility study involving a Trappist beer. *Journal of the Institute of Brewing,* **116,** 56-61.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta,* **653,** 121-130.

Williams, P. J., Geladi, P., Britz, T. J. & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science,* **55,** 272-278.

Williams, P. J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**, 131-138.

Yoshizawa, T., Yamashita, A. & Luo, Y. (1994). Fumonisin occurrence in corn from high-and low-risk areas for human esophageal cancer in China. *Applied and Environmental Microbiology*, **60**, 1626-1629.

Zhu, L. J., Dogan, H., Gajula, H., Gu, M.-H., Liu, Q. Q. & Shi, Y. C. (2012). Study of kernel structure of high-amylose and wild-type rice by X-ray microtomography and SEM. *Journal of Cereal Science*, **55**, 1-5.

CHAPTER 2
LITERATURE REVIEW

Current maize grading is a tedious manual process that could be improved by utilising modern analytical techniques. Other manual procedures, for example used in cereal research, have further encouraged the development of non-destructive, rapid and accurate analytical techniques to evaluate the grain quality and safety. Near infrared (NIR) hyperspectral imaging integrates spectroscopy and imaging techniques in one analytical system, allowing the identification of chemical components and their spatial distribution within the sample. Individual kernels can be analysed non-destructively, enabling one to follow changes in the same sample over time. NIR hyperspectral imaging has gained popularity for acquiring rapid chemical and physical information. It is capable of quantifying, identifying or differentiating a variety of properties. Although NIR hyperspectral imaging has not yet been extensively implemented in industry, it shows great potential in laboratory studies for the development of an on-line evaluation system to assess cereals. Industry could soon move away from subjective, manual classification and measuring methods. NIR hyperspectral imaging methods could replace multiple conventional chemical, microbial or physical tests with a single, automated image acquisition. This review outlines the current status of maize grading in South Africa, as well as theory, principles and applications of NIR hyperspectral imaging in cereal quality and safety evaluation.

Accepted for publication:

Sendin, K., Williams, P. J. & Manley, M. (2016). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical reviews in Food Science and Nutrition*.

DOI: 10.1080/10408398.2016.1205548

INTRODUCTION

Cereals and cereal products form a major part of the global population's diet, where wheat, barley, oat, rye, maize, sorghum and millet provide 56% of the food energy consumed on earth (Koohafkan and Stewart, 2008). Wheat and rice are by far the most widely consumed cereals, while maize plays a major role directly as a human food source, and indirectly as animal feed. Together these three account for 85% of the overall global cereal production. The processing ability and end-product quality of cereal products are highly influenced by the grade of the grain used as the raw material. Furthermore, it is the basis for allocating a market price. The quality evaluation of cereals considers multiple quality factors such as chemical and physical properties, growing region, growing season, colour, texture, protein content, and kernel hardness. Various safety parameters are also considered, such as microbial, parasitic, chemical and physical contamination, and several specific defects. Varietal identification is also important, not necessarily for food manufacturers, but more specifically for plant breeders to differentiate between genotypes.

In the foreseeable future, the cereal industry will increasingly move away from manual methods and toward rapid, non-destructive, non-invasive and accurate analytical techniques. Particularly applicable are techniques to evaluate and differentiate between kernels for the purpose of grading and milling quality assessment. Studies have demonstrated the limited human capacity to reproduce a consistent estimation of quality (Lorente *et al.*, 2012). It was also shown that, in processes like cereal grading, this inconsistency increases when a number of properties must be evaluated simultaneously. Ideally techniques should have the potential to be implemented on-line in a mill or factory. Near infrared (NIR) hyperspectral imaging is a technique that meets all of these requirements. As a consequence, it gained popularity in research for identifying, differentiating or evaluating various properties of cereals, such as hardness classification (Gorretta *et al.*, 2006; Manley *et al.*, 2011; Manley *et al.*, 2009; Williams *et al.*, 2009; McGoverin and Manley, 2012; Mahesh *et al.*, 2015), chemical composition (Cogdill *et al.*, 2004; Lin *et al.*, 2006; Weinstock *et al.*, 2006; Mahesh *et al.*, 2015), variety identification (Choudhary *et al.*, 2009; Mahesh *et al.*, 2008; Wang *et al.*, 2014a; Wang *et al.*, 2015), sprouting detection (Arngren *et al.*, 2011; McGoverin *et al.*, 2011; Singh *et al.*, 2009b), physical quality classification (Archibald *et al.*, 1998; Wallays *et al.*, 2009), fungal contamination detection (Bauriegel *et al.*, 2011; Berman *et al.*, 2007; Del Fiore *et al.*, 2010; Vermeulen *et al.*, 2012; Wang *et al.*, 2014b; Williams *et al.*, 2012; Yang *et al.*, 2012), and parasitic contamination detection (Singh *et al.*, 2009a; Singh *et al.*, 2010).

Conventional NIR spectroscopy is popular and well established, in cereal research and industry as well as throughout food science and technology applications, as a non-destructive and rapid analytical technique (Manley, 2014). Hyperspectral instruments operating in the NIR region show the greatest potential for the study of heterogeneous biological or agricultural materials, such as cereals. This is due to the fact that the NIR region characteristically excites major X–H bonds, namely C–H, O–H and N–H, common constituents of organic molecules.

The early discovery of the NIR region (780–2500 nm) by Frederick William Hershel in 1800 was followed by a long period in which his findings were not utilised in science and research (Manley, 2014). The development of NIR technology was later revived by Karl Norris and his co-workers, whose scientific work gained momentum in the 1960s, and focused on the agricultural sector (Dale *et al.*, 2013). The first notable use of NIR spectroscopy in cereal science was the work of Williams (1975), who measured the protein and moisture content of wheat as a quality basis for wheat trading. In recent years, new techniques and methods based on NIR spectroscopy have been developed. There are other notable uses of NIR spectroscopy in combination with other methods, such as NIR microscopy (Yang *et al.*, 2012). The integration of NIR spectroscopy technology with imaging techniques led to the development of NIR hyperspectral imaging, which has found application in a large number of diverse fields.

The basis of spectroscopy is the interaction between incident light and molecules, where different molecules are sensitive to light of different wavelengths (Feng and Sun, 2012). The resultant spectra reveal how light was absorbed or scattered, thus revealing information on the sample and its constituents. This information may be physical, chemical, or biological. While this review focuses mainly on studies in the NIR region, spectroscopic techniques utilise a number of regions in the electromagnetic spectrum. By exciting molecules with light, electrons can move through various energy states (Li-Chan, 2011). These states include levels known as electronic states. Within one electronic state level exists numerous vibrational and rotational states. Excitation with the longer wavelength (and thus lower energy) NIR region allows an electron to only move between vibrational states within a single electronic state. When higher energy ultraviolet (UV) and visible radiation is used, electrons will begin to move from one electronic state to another. The spectrum resulting from a vibrational technique, such as NIR spectroscopy or spectral imaging, generally exhibits broad peaks. Due to multiple vibrational states, the individual bond stretches and bends are coupled. The resulting absorptions are combinations of the bending and stretching of several bonds. Due to the nature of the electronic states occupied when using UV and visible wavelengths, peaks tend to be sharper and more defined.

One constraint of conventional spectroscopy methods is that only a small portion of the sample is scanned and an average spectrum or prediction value for all kernels in a bulk sample is given. It thus may not be representative of the entire sample or sample group, especially if it is heterogeneous. In order to spatially resolve this spectral information, computer vision is utilised. Thus, one can identify individual kernels out of a sample group, and also observe spatial distributions (e.g. distribution of moisture) in a single kernel. Computer vision gathers information on the shape, size, texture, as well as colour, but the usual three bands (red, green and blue) used are often not sufficient for detecting chemical and biological properties (Feng and Sun, 2012). Thus, NIR hyperspectral imaging is a natural expansion of conventional spectroscopy and computer vision to provide spatial positional information of the acquired spectra (Burger and Geladi, 2006).

This review addresses the current requirements of the maize industry regarding grading legislation for human consumption and the possible utilisation of NIR hyperspectral imaging for grading. The history, principles and theoretical background of NIR hyperspectral imaging are explored and explained. The different hyperspectral image acquisition configurations are discussed and the multivariate techniques most commonly used are reviewed. The numerous applications of this technique in quality and safety control of cereals and cereal products are discussed to illustrate the potential NIR hyperspectral imaging holds for the future of the maize industry.

MAIZE GRADING IN SOUTH AFRICA

A very simple manual grading method is followed (Department of Agriculture, 2009). A sample with a minimum weight of 150 g (*ca.* 600 kernels) from a maize consignment is visually inspected to identify any content other than sound white maize, such as defective or foreign material. The method involves spreading the kernels onto a flat surface and slowly inspecting each kernel. The kernels are turned over to ensure both sides are sound. Any undesirable materials are weighed and referenced to maximum levels stipulated for each material's category. There are three white maize classes, WM1 (best), WM2 and WM3. The grade is allocated solely on the contents of undesirable materials within a sample, as given in Table 2.1. Refer to Addendum 1 for further legislative details. Other cereal commodities, such as wheat, entail more elaborate grading systems and include steps additional to visual inspection, such as hectolitre mass, falling number, moisture content and protein content measurements (Department of Agriculture, 2016). It is always essential to obtain a completely random and representative 150 g sample, as the outcome determines the grade of an entire truckload, railcar, or sub-lot of a shipping vessel.

The South African legislation identifies four main categories of undesirable materials, namely defective maize, other colour maize, pinked maize and foreign matter. Within these main categories are sub-categories. Defective maize kernels most commonly occurring in South Africa include heat damage, water damage, screenings/broken kernels, and fungal damage. Other defects, such as frost damage or sprouting, are included in the legislation, however are very rare in the South African industry. Some of these defects not only affect the quality characteristics of the maize, but also pose potential safety risks. Thus, this main category has a maximum levels of 7% for WM1. The other colour maize category includes only yellow maize, as only white and yellow maize are grown commercially in South Africa. Yellow maize can cause overall discolouration in white maize products, and thus levels are limited to 3% for WM1. The pinked maize category includes only pink discoloured white maize, a defect which poses no safety threat and does not cause severe colour changes, and is thus leniently restricted (20% for WM1). Lastly, foreign matter is divided into two main categories. Foreign materials that pose a safety risk, such as glass, stone, coal, dung or

Table 2.1 Standards for grades of white maize (per category of undesirable material and maximum allowed weight) according to South African grading regulations.

Defect class	Category	WM1	WM2	WM3
		(%)	(%)	(%)
***Fusarium* fungal damage** ***Diplodia* fungal damage** **Water damage** **Heat damage** **Rodent damage** **Screenings**	Defective white maize kernels	7	13	30
Plant material **Wheat** **Sorghum** **Soy beans** **Sunflower seeds**	Foreign matter	0.3	0.5	0.75
Yellow maize	Other colour kernels	3	6	10
Pinked maize	Pinked white maize kernels	20	20	20
WM = White maize grade				

metal, must be completely absent in all grades. These materials are removed during a preliminary dockage sorting step and are unlikely to be found in the grading room. Less dangerous foreign materials, for example other cereal commodities (wheat, soy, sunflower seeds and sorghum) or maize plant materials, are allowed within the given limits (0.3% for WM1).

Heat damage can occur under excessive artificial drying temperatures or poor storage conditions (Symons and Shahin, 2008). These kernels are identified by a scorched dark yellow discolouration. The damage most commonly occurs when maize is stored at high humidity, causing high respiration rates. Intrinsic enzymes are activated, causing soluble sugars to be formed (Serna-Saldivar, 2010). Due to the high heat of respiration and presence of reducing sugars, Maillard reactions and disclouration occur. These kernels contain high amounts of degraded starch which affect milling and processing ability.

Water damage occurs when maize is exposed to wet conditions pre-harvest. As the kernels are still attached to the maize cob when damage occurs, the defect is characterised by the development of a dark ring on the lower section of the kernel near the point of attachment. Water damage can predispose maize to a risk of further damage and infestations.

Screenings are measured using a screen or sieve to separate whole kernels from smaller broken pieces, hence their naming. This sub-category includes any fine maize particles and other

materials that are able to pass through a 6.35 mm diameter sieve. Harvesting introduces substantial mechanical damage to kernels, which is further aggravated by subsequent handling and transportation operations. Such damage includes hairline cracks, and broken, chipped or crushed kernels, resulting in a shorter storage life (Symons and Shahin, 2008).

Numerous pests and parasites contaminate and damage maize pre- and postharvest, as well as during the storage period. Rodent and insect damage is generally considered the leading cause of grain losses, where rodent damage is substantially more prevalent in South Africa. Rodents characteristically eat only the germ of kernels, while the rest of the kernel remains intact (Mdangi *et al.*, 2013). The nutritional content is reduced and, furthermore, contamination of the maize with faeces, hair and urine can cause additional safety issues.

Fungal infestations cause a significant amount of deterioration, and are the second largest source of maize losses (Ominski *et al.*, 1994). Fungal contamination not only leaves grain unfit for food production by discoloration and reduction of nutritional value, but often results in mycotoxin production (Fandohan *et al.*, 2004). Fungi can infest maize in the field and during storage, provided that the moisture content is above the critical level of 14% (Serna-Saldivar, 2010). Fungal infestations are usually recognised by colour changes on the pericarp and germ tissues. Amylases, lipases and proteases break down starch, lipids and proteins, respectively, leading to lower test weight and high contents of reducing sugars and fat rancidity. Although graders utilise visual inspection to identify the fungal damaged kernels, severely affected samples will also have a characteristic odour. The two fungal diseases occurring widely throughout South African maize are *Fusarium* and *Diplodia* infections. *Fusarium* species were found to contaminate as much as half of all maize before harvest (Fandohan *et al.*, 2004). It is an aggressive invader that destroys starch granules, storage proteins and cell walls, and generally leaves kernels shrivelled and discoloured to brown or purple. Studies on the occurrence of oesophageal cancer showed a link to the consumption of *Fusarium* contaminated maize, due to the fumonisin mycotoxins (Sydenham *et al.*, 1990; Yoshizawa *et al.*, 1994). *Diplodia*, often also referred to as *Stenocarpella*, a common contaminant of maize in tropical and sub-tropical regions can result in significant yield losses with lightweight, and brown discoloured and shrivelled kernels. In addition, *Diplodia* is known to produce mycotoxins (Masango *et al.*, 2014).

Any potential analytical technique investigated for automating maize grading must be capable of accurately evaluating these four categories of undesirable materials. These materials are diverse, and past applications of NIR spectral imaging in cereal research must be reviewed to demonstrate that the technique can deliver high accuracy in evaluating a wide variety of cereal properties.

BRIEF HISTORY OF HYPERSPECTRAL IMAGING

Hyperspectral imaging has transformed through the years from when the technique was first used in the field of remote sensing (Goetz *et al.*, 1985). A key development in the technology was the invention by George Smith and Willard Boyle in 1979; the first hybrid area array detectors using mercury-cadmium-telluride on silicon charged-coupled devices (CCD). These detectors allow a camera to record both spatial and spectral information of wavelengths beyond 1.0 μm (Thenkabail, 2012). Hyperspectral imaging was initially used to image and identify surface materials for geological purposes, such as airborne mineral mapping. It was particularly useful as the device could 'see through' vegetation covering the land.

The growth of hyperspectral imaging required developments in electronics, hardware, computing, and software throughout the 1980s and into the 1990s before the scientific community would begin to embrace the technique (Goetz, 2009). Much of the hardware development took place at the National Aeronautics and Space Administration Jet Propulsion Laboratory (NASA/JPL), leading to one of the most notable early uses of hyperspectral imaging, the Airborne Visible/Infrared Imaging Spectrophotometer (AVRIS) in 1987. Since the success of AVRIS, developments in sensor, calibration and data systems gained momentum, and scientists began to incorporate multi- and hyperspectral instruments in many laboratory, industrial and airborne applications (Elmasry *et al.*, 2012). In recent years, hyperspectral imaging has been used in diverse fields such as astronomy (Hege *et al.*, 2004), security and defence (Patterson and Brescia, 2010), precision agriculture (Nansen *et al.*, 2015; Jay *et al.*, 2010), environmental monitoring (Keith *et al.*, 2009; Spangler *et al.*, 2010), geology (Resmini *et al.*, 1997), pharmaceuticals (Ravn *et al.*, 2008), biotechnology (Polerecky *et al.*, 2009) and food quality and safety (Elmasry *et al.*, 2011; Zhu *et al.*, 2013; Nansen *et al.*, 2008).

PRINCIPLES AND THEORETICAL BACKGROUND OF HYPERSPECTRAL IMAGING

Fundamentals of Hyperspectral Imaging

Through the integration of spectroscopy and computer vision, one is provided with both the chemical composition and the physical properties of the sample (Feng and Sun, 2012). Thus, for the first time in scientific history, researchers have a non-destructive analytical tool that provides spatially resolved chemical information about biological and agricultural materials (Baeten *et al.*, 2007).

Hyperspectral imaging involves the simultaneous acquisition of spatial images in numerous spectrally congruent bands (Burger and Geladi, 2006). The hyperspectral image formed is a stack of hundreds of two-dimensional spatial image planes of the same sample. Each plane comprises its own narrow band of wavelengths, thus, by stacking the layers, a complete spectrum for each pixel

is formed. This collection of spectral layers, arranged in a three-way data matrix, is termed a hypercube, or alternatively spectral cube, data cube, or spectral volume (Manley, 2014). Pixels are digitised grey values (intensities) at a certain wavelength and are expressed as integers. These intensity values generally have 8-, 12-, 14- or 16-bit grey values, where 12-bit is generally adequate for most applications (Elmasry and Sun, 2010). The hypercube consists of grey value measurements in three dimensions, namely two spatial dimensions (x rows $\times$ y columns), and one spectral dimension (z wavelengths) (Dale *et al.*, 2013), as illustrated in Fig. 2.1. This allows one to either view the hyperspectral image as an image plane taken at a single wavelength, or as a spectrum for a particular pixel.

The underlying principle of hyperspectral imaging, and all spectroscopic techniques in fact, is that all materials, due to their inherent chemical composition and physical structure, will reflect, scatter, absorb or emit electromagnetic radiation in characteristic patterns at specific wavelengths. The spectrum, or spectral signature, of a sample is generated by plotting light intensity against the wavelengths recorded. The spectral signature of the sample is easily used to classify, detect or differentiate the given material(s) in every pixel of the hypercube. Peaks at specific wavelengths give an indication of the associated chemical compounds. With the appropriate data processing, one may routinely categorise the location of constituents that display particular spectral signatures, and thus map their distribution within the sample, making the analysis ideal for heterogeneous

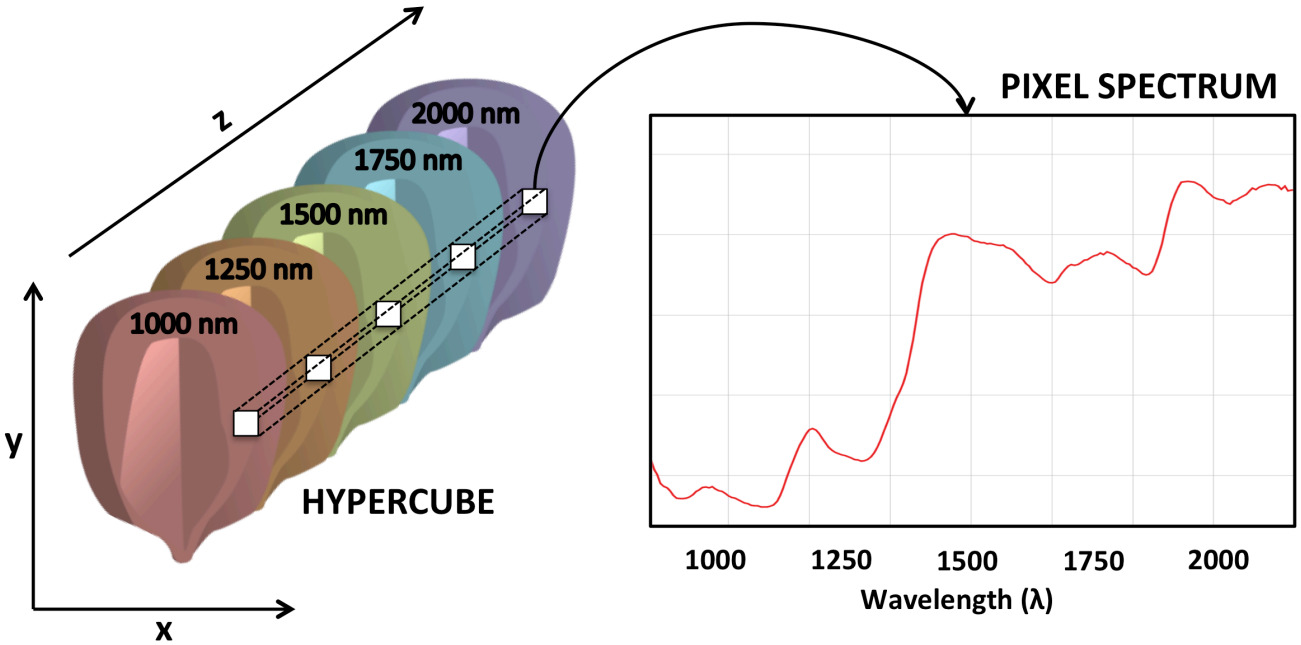

Figure 2.1 A NIR hyperspectral imaging hypercube comprises of one wavelength (z) and two spatial (x and y) dimensions. One can view the data as a spectrum of a pixel in the sample, or as an image plane of the entire sample at a chosen wavelength. (e.g. at 1000 nm).

samples (Elmasry *et al.*, 2012; Manley, 2014). NIR hyperspectral imaging utilises a spectral range of 780 to 2500 nm. Spectra from this region contain information regarding major chemical constituents of organic molecules, namely C–H, O–H and N–H bonds (Manley, 2014), and thus hyperspectral imaging in the NIR region is an excellent tool for studying biological materials.

Components of Hyperspectral Imaging Systems

The more popular approach used to generate hyperspectral images, especially in food inspection, is the pushbroom method. A schematic representation of the configuration is given in Fig. 2.2. A typical hyperspectral imaging instrument comprises an illumination source; an objective lens; a spectrograph; a camera; and a computer equipped with image acquisition software (Elmasry *et al.*, 2012; Gowen *et al.*, 2007).

The spectral range of the instrument will be determined by the camera, illumination source, and detector used (Gowen *et al.*, 2007). Instruments operating in the Vis-NIR range of 400 to 1000 nm utilise lower cost CCD or complementary metal oxide semiconductor (CMOS) silicon-based detectors to collect both spectral and spatial information (Gowen *et al.*, 2007; Manley, 2014). If NIR wavelengths of 1100 nm to 2500 nm are required, an expensive indium gallium arsenide (InGaAs) or mercury cadmium telluride (HgCdTe) array detector is used. Tungsten-halogen lamps are the most common illumination sources used in NIR hyperspectral imaging (Elmasry and Sun, 2010).

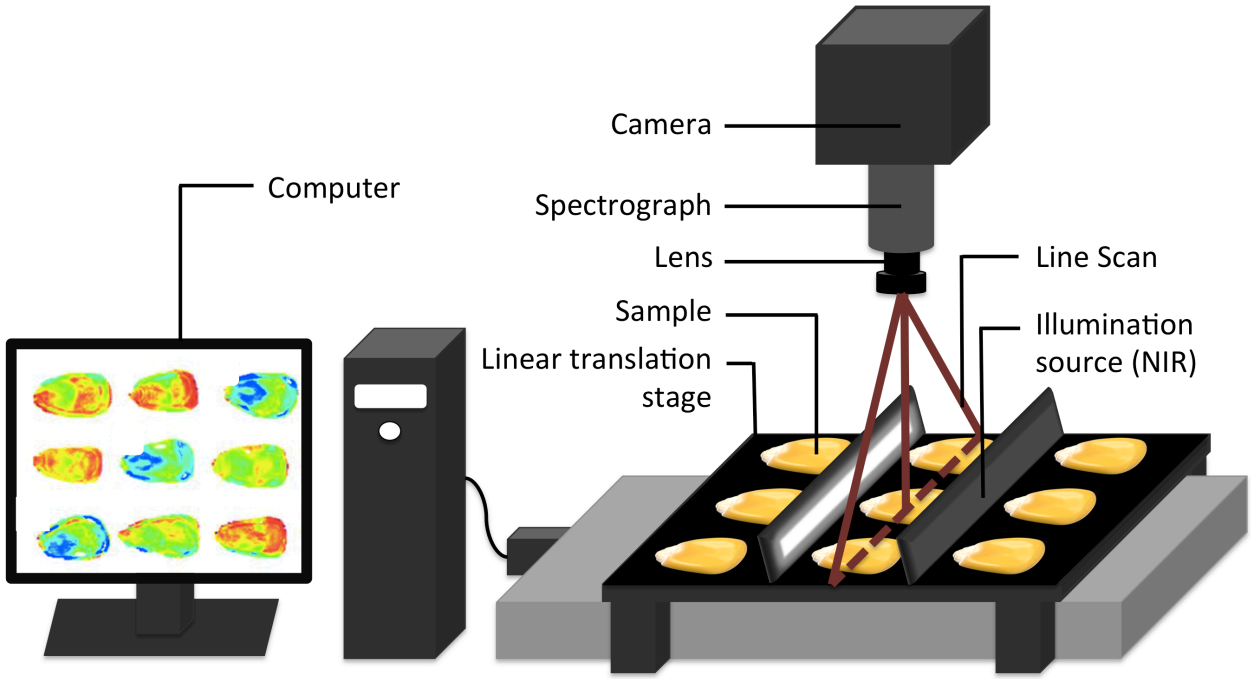

Figure 2.2 Schematic of a pushbroom NIR hyperspectral imaging system, illustrating the line-by-line data acquisition of the samples on the linear translation stage.

This is due to their durability, stability, and capability to emit light of a broad spectral range (400 – 2500 nm). Other illumination sources include quartz-halogen lamps, light emitting diodes (LEDs), tunable lasers, and heated xenon lamps. LED sources are, however, restricted to a narrow range of 300 to 1000 nm.

The data collection and analysis require powerful computers and significant data storage capacity, as the collected image data are large, often further contributing to the already high cost of hyperspectral instruments (Manley, 2014).

Acquisition of Hyperspectral Images

Two main instrumentation approaches are used for the generation of hyperspectral images, namely the (a) pushbroom or line scan; and (b) staring scan (Dale *et al.*, 2013). Both of these methods acquire information from a series of two-dimensional images to form the final three-dimensional hypercube. The pushbroom instrument captures all spectral information simultaneously across one spatial dimension, and gradually acquires the second spatial dimension line-by-line. The staring scan instrument collects all data points in both spatial dimensions simultaneously, and gradually acquires the spectral dimension, one waveband at a time. It is currently impossible to acquire information in all three dimensions simultaneously (Gowen *et al.*, 2007).

In a pushbroom system (Fig. 2.2), the instrument moves the sample through the projected broadband beam of light (Lorente *et al.*, 2012). The light interacts with the sample and a smaller amount is reflected back. The reflected light passes through the objective lens, and is then collimated, dispersed by a grating into different directions according to wavelengths, and focused on the detector. The two-dimensional plane (one spatial and one spectral dimension) formed of the scanned line is captured by the camera and stored on the computer. To generate the second spatial dimension, the sample is placed on a linear translation stage, which is slowly moved across the objective lens to acquire successive planes. These are compiled to form the complete hypercube. Advances in detector capability have allowed this method to become very rapid, enabling this system to be implemented on-line in industry as samples move on a conveyer belt.

The staring scan instrument uses a filter to select one wavelength at a time and consecutively acquire a spatial image of the entire sample as the instrument successively passes through the selected spectrum (Manley, 2014). Despite the difference in the order in which the data is collected, the spectroscopic principles of the staring scan instrument are similar to the pushbroom configuration. The most frequently used filters are liquid crystal tunable filters (LCTF) and, to a lesser extent, acousto-optic tunable filters (AOTF). High spatial resolution is possible with staring scan image acquisition, allowing this technique to be paired with microscopy (Dale *et al.*, 2013). However, a disadvantage of this technique is that changes may take place in the sample during scanning, as the scan may require several minutes. Furthermore, during this time, the heat generated from the lamp may cause changes in the sample.

Analysis of Hyperspectral Images

Analysing the huge data sets generated from hyperspectral image acquisition has been one of the technique's most prominent challenges in almost all fields of application. The main issue is being able to extract only the meaningful information from the raw data. Chemometrics is used to reduce the size of the data set while retaining essential spectral information, and to identify the important characteristics of an image. Multivariate data analysis is the chemometric tool used, and is the key to correlating features of the hyperspectral image to the properties of the sample being studied. Cereal science applications usually use principal component analysis (PCA) as the method of unsupervised qualitative and exploratory multivariate data analysis (Manley *et al.*, 2011; Williams *et al.*, 2009; Weinstock *et al.*, 2006; Choudhary *et al.*, 2009; Wang *et al.*, 2014a; Arngren *et al.*, 2011; McGoverin *et al.*, 2011; Singh *et al.*, 2009a; Singh *et al.*, 2009b; Archibald *et al.*, 1998; Bauriegel *et al.*, 2011; Del Fiore *et al.*, 2010; Wang *et al.*, 2014b; Williams *et al.*, 2012; Yang *et al.*, 2012). Minimum noise fraction (MNF) transformation (Arngren *et al.*, 2011) is another unsupervised technique which is essentially two cascaded PCA transformations. Partial least squares (PLS) (Cogdill *et al.*, 2004), PLS discriminant analysis (PLS-DA) (Cogdill *et al.*, 2004; Del Fiore *et al.*, 2010; Manley *et al.*, 2009; McGoverin *et al.*, 2011; Vermeulen *et al.*, 2012; Wallays *et al.*, 2009; Williams *et al.*, 2009) and linear discriminant analysis (LDA) (Singh *et al.*, 2009a; Singh *et al.*, 2009b; Singh *et al.*, 2010; Yang *et al.*, 2012) are used for supervised classification, while PLS regression (PLSR) is often used for quantification, as well as supervised classification (Lin *et al.*, 2006; Weinstock *et al.*, 2006; Williams *et al.*, 2012).

Data analysis can be broken down into two stages, hypercube pre-processing and hypercube processing. Pre-processing of both spectral and spatial data reduces the data set size and removes the influence of undesirable phenomena, such as the background surface data (e.g. petri dish), light scattering and shadows, particle size effects or detector artefacts (Vidal and Amigo, 2012). Image compression is performed to reduce the sometimes millions of data points to a more manageable data set, retaining only essential data. For this wavelengths known to be associated with the properties studied may be selected, or a method such as PCA is utilised. The area of interest is selected using a technique called brushing (Esbensen and Geladi, 1989), whereby the often noisy image background data is eliminated. Dead pixels, pixels presented as zero values due to detector anomalies, should be detected and removed. Lastly, spectral pre-processing is performed. Scattering effects and undesirable spectral variations must be suppressed, where techniques such as standard normal variate (SNV) (Barnes *et al.*, 1989) or multiplicative scatter correction (MSC) (Esbensen and Geladi, 1989) may be applied to remove scattering and smooth noisy spectra. One may also choose to highlight spectral features using derivatives. Collectively, these pre-processing techniques will help extract information, but can reduce spatial and spectral resolution, and must thus be performed with consideration of the study's aim.

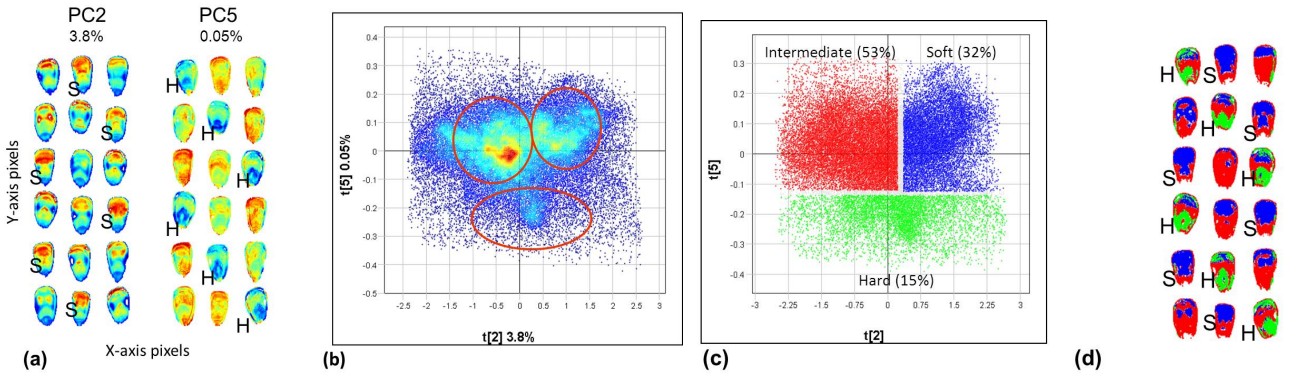

Figure 2.3 (a) Scores images of PC2 and PC5 for whole yellow maize kernels enabling visualisation of similarity in chemical composition (similar colours indicate similar chemical composition, in this case similar endosperm texture). (b) Scores plot of PC2 vs. PC5 with three clusters. (c) Classification plot based on clusters identified in the PC scores plot. (d) Classification image after projection of the classes identified in the scores plot onto the scores image [Copyright 2014 Royal Society of Chemistry. Reproduced with permission from Manley (2014)].

PCA is one of the most versatile and widely applied techniques used before calculating any other multivariate model (Esbensen *et al.*, 2002; Cowe and McNicol, 1985). It is a robust and powerful unsupervised technique, and is the initial step in data processing to provide insight and reduce the dimensionality of data for building more complex models. PCA provides an overview of the information in the dataset by identifying the main sources of variation within and between samples. New variables are formed as uncorrelated and linear combinations of the original variables (Lorente *et al.*, 2012). A few of these new variables, called principal components (PCs), ideally explain the majority of the common variations within the data. However, as PCA is an unsupervised classification technique, PCs are not calculated based on the defined class segregations of the experiment. The variance identified in the first PCs may not necessarily be related to the classes or properties under investigation. Relevant information is often observed in lower-order PCs, and it is thus essential to also evaluate these PCs as carefully as the initial few (Manley, 2014).

Results may be visualised using PCA (Manley, 2014), as illustrated in Fig. 2.3. The similarity in samples is displayed using a heat map, where similar colours represent similar score values (Fig. 2.3a). This similarity in scores value is often due to similar characteristics e.g. chemical composition. Similarity between pixel spectra is shown as clusters on the PC scores plot (Fig. 2.3b). A classification plot may be generated by assigning each cluster to a class (Fig. 2.3c). When these assigned class's dummy variables are projected onto the scores image, a classification image is generated (Fig. 2.3d). In this example, the clusters in the PCA scores plot obtained were due to differences in endosperm texture of whole maize kernels.

In discrimination studies, supervised classification usually follows, often using LDA or PLS-DA (Amigo *et al.*, 2013). Discriminant analysis classifies objects into mutually exclusive classes

based on a set of measurable features of the objects, normally spectral features arising from differences in the mean spectra of the classes (Lorente *et al.*, 2012). Thus, only a sub-set of the hyperspectral image data that maximises differentiation is used. The independent variable is a categorical one, and defines the class membership of the samples. LDA calculates an optimal linear projection that maximises the variance between the classes, while keeping the variance within a class to a minimum (Fisher, 1936). All of the data may then be projected accordingly. Pixels or objects (depending on the nature of the study) are classified by calculating their distance to the centre of each class, and assigning them to the class with the smallest distance. PLS-DA differs in that a threshold is chosen with a value between the average spectra of each of the classes (Chevallier *et al.*, 2006). A threshold of 0.5 is often used, but the best fitting value depends on the ratio of number of pixels or objects in each of the classes and may be chosen by trial and error. The predicted values of pixels or objects above or below the threshold will be classified as the different classes. These two techniques can be used to produce monochromatic classification images revealing how pixels or objects are classified. PLS-DA generally performs better when the classes are very similar, as it is able to overcome the collinearity problems associated with LDA.

To obtain quantitative information, PLS regression is often used. The predictor (spectrum) is the X-variable, and the response to be modelled (e.g. concentration) is the y-variable (Næs *et al.*, 2002). First a sample set with a range of known **y**-values is imaged. The representative spectrum of each sample is pooled to form a matrix (**X**), which represents the corresponding y-values for each pixel in the image. This relationship is modelled by estimating a regression vector, **b**. To estimate this vector, PLS regression identifies components that are aligned with maximum variance in **X**, similarly to PCA, but also considers the covariance with **y**. The correct number of components included in the calibration model must be chosen to ensure sufficient prediction accuracy, giving the regression vector. If too many are chosen, the model becomes over-fitted, and the parameters are too strictly defined, leading to a less robust model. Using this model, one can subsequently predict the unknown **y**-variable of a sample. If quantitative information is not required, PLS regression can also be used for classification, as used by Williams *et al.* (2012). The pixels of an image will simply be compared amongst themselves, as opposed to being assigned quantitative y-values. The values of pixels themselves have no specific unit of measurement, but are assigned relative values ranging between 0 and 1. The PLS regression results vary from the PLS-DA results as they are classified as continuous data, and not as categorical or binary data (only 0 or 1). Thus, a classification image from PLS regression will be displayed in a colour scale, instead of the monochrome images from PLS discriminant analysis.

Other multivariate data analysis tools include soft independent modeling of class analogy (SIMCA) for unsupervised classification; artificial neural networks (ANN) for unsupervised or supervised classification and quantification (Mahesh *et al.*, 2008; Lin *et al.*, 2006); spectral angle mapping (SAM) for unsupervised classification (Bauriegel *et al.*, 2011); multiple linear regression

(MLR) for supervised quantification (Lin *et al.*, 2006); factorial discriminant analysis (FDA) for supervised classification (Wang *et al.*, 2014b; Gorretta *et al.*, 2006); quadratic discriminant analysis (QDA) for supervised classification (Mahesh *et al.*, 2008; Choudhary *et al.*, 2009; Singh *et al.*, 2009a; Singh *et al.*, 2009b); penalised discriminant analysis (PDA) for supervised classification (Berman *et al.*, 2007); support vectors machine (SVM) for unsupervised pattern recognition, supervised classification and quantification (Vermeulen *et al.*, 2012); back propagation neural network (BPNN) for supervised classification (Choudhary *et al.*, 2009; Wang *et al.*, 2014a; Singh *et al.*, 2010); linear discriminant classifier (LDC) for supervised classification (Choudhary *et al.*, 2009); minimum noise fraction (MNF) for supervised classification (Arngren *et al.*, 2011); and principal component regression (PCR) for quantification (Cogdill *et al.*, 2004).

APPLICATIONS OF HYPERSPECTRAL IMAGING IN CEREAL EVALUATION

The applications of NIR hyperspectral imaging in cereal science are numerous, and may be broadly divided into quality and safety. The general information of all studies reviewed is given in Table 2.2.

Quality

Hardness Classification

The hardness of grain is determined by the content of vitreous and floury endosperm in the kernel. Vitreous endosperm has a stronger protein matrix and leads to harder grain, while floury endosperm has a higher starch content and is less compact. Hardness is an indicator of the kernel's density, bulk density, and breakage susceptibility (Williams *et al.*, 2009). As hardness influences end-use processing performance, its determination is an important tool in the cereal industry. In wheat, the terms hard and soft generally describe a variety's underlying genetics. When testing, the hardness measurement will measure the way in which the endosperm fractures during milling. Wheat will thus be classified as soft (biscuit wheat) and hard (bread wheat), as well as very hard (durum) (Giroux and Morris, 1997). Maize may contain both vitreous and floury sections of endosperm, and hardness is classified by the ratio of the two as soft, intermediate and hard (Manley *et al.*, 2009).

NIR hyperspectral imaging has been used to differentiate between vitreous and floury endosperm in yellow maize kernels (Williams *et al.*, 2009; Manley *et al.*, 2009). Both studies developed PLS-DA models that performed well. Williams *et al.* (2009) imaged datasets of different sizes using two different cameras, resulting in three models with coefficient of determination (R^2) and root mean square error of prediction (RMSEP) values ranging from 0.76 to 0.88 and 0.18 to 0.29, respectively. Even with the small test sets of 12 and 24 kernels, the results were reproducible between data sets and between instruments, indicating that NIR hyperspectral imaging has a real potential for future classification uses. Furthermore, other typically used maize kernel hardness tests (e.g. particle size index, Stenvert test, test weight (Blandino *et al.*, 2010; Fox and Manley, 2009)) were not capable of differentiating between the two endosperm types within the kernels.

Table 2.2 Applications of hyperspectral imaging and chemometrics for quality and safety evaluation of cereals and cereal products.

APPLICATION	IMPORTANT WAVELENGTHS (nm)	CHEMOMETRICS	R^2	REFERENCE
Maize				
Hardness prediction	1400	PCA*; PLS-DA	0.85; 0.88; 0.76	(Williams *et al.*, 2009)
Hardness prediction	1220; 1405; 1690; 1870	PCA*; PLS-DA	0.85	(Manley *et al.*, 2009)
Hardness prediction	1367; 1411; 1455; 1925; 1974; 2000; 2148	PCA*	-	(McGoverin and Manley, 2012)
Moisture content prediction	910; 960-970	PLS; PCR	0.87	(Cogdill *et al.*, 2004)
Oil content prediction	910; 960-970	PLS; PCR	0.85	(Cogdill *et al.*, 2004)
Oil content prediction	-	PCA*; PLSR	0.68	(Weinstock *et al.*, 2006)
Oleic Acid content prediction	-	PCA*; PLSR	0.65	(Weinstock *et al.*, 2006)
Variety identification	438; 487; 511; 516; 700; 897	PCA; SVM	89; 90%**	(Wang *et al.*, 2015)
Aspergillus and *Fusarium* detection	410; 470; 535; 945	PCA; PLS-DA	-	(Del Fiore *et al.*, 2010)
Fusarium detection	1405; 1660–1668; 1900; 2136	PCA*; PLSR	0.71; 0.63	(Williams *et al.*, 2012)
Wheat				
Hardness prediction	910; 990; 1030	PLS-FDA	94%**	(Gorretta *et al.*, 2006)
Hardness prediction	1180–1220; 1320–1400; 1460–1490	PLSR; PCR	0.82; 0.75	(Mahesh *et al.*, 2015)
Hardness effects	1430; 1886; 1899; 1920; 1936; 1940; 1984	PCA*	-	(Manley *et al.*, 2011)
Protein content prediction	960–1030; 1180– 1200; 1460–1500; 1670–1700	PLSR; PCR	0.68; 0.62	(Mahesh *et al.*, 2015)
Variety identification	1130; 1310	LDA; QDA; ANN	>94; >86; >80%**	(Mahesh *et al.*, 2008)
Variety identification	960; 970; 980; 1160; 1680; 1700; 1450	LDC, QDA; BPNN	99; 98; 92%**	(Choudhary *et al.*, 2009)
Sprouting detection	1102; 1132; 1305	PCA*; LDA; QDA; Mahalanobis DA	94%; 94%; 95%**	(Singh *et al.*, 2009b)

Sprouting detection	1920–1940	PCA*; PLS-DA	0.67	(McGoverin *et al.*, 2011)
Colour classification	990; 1024	PCA	-	(Archibald *et al.*, 1998)
Fusarium detection	500–533; 560–675; 682–733; 927–931	PCA*; SAM	87%**	(Bauriegel *et al.*, 2011)
Fungal contamination detection	568; 900; 1100; 1300; 1000; 1200	PDA	95%**	(Berman *et al.*, 2007)
Aflatoxin B_1 detection	459; 1706; 1729; 1934; 2057; 2344; 2384	PCA*; FDA	94%**	(Wang *et al.*, 2014b)
Ergot body detection	-	PLS-DA; SVM	0.99	(Vermeulen *et al.*, 2012)
Insect contamination detection	1102; 1305	PCA*; LDA; QDA	85–100%**	(Singh *et al.*, 2009a)
Insect contamination detection	870	PCA*; LDA; QDA; Mahalanobis DA	79; 86; 74%**	(Singh *et al.*, 2010)
Rice				
Moisture content prediction	960; 930; 980; 1000; 940; 1014	MLR; PLSR; ANN	0.94; 0.95; 0.95	(Lin *et al.*, 2006)
Variety identification	419; 452; 593; 613; 743; 784; 966	PCA; BPNN	89; 94%**	(Wang *et al.*, 2014a)
Pyricularia detection	1188; 1339; 1377; 1432; 1614	PCA*; LDA	92%**	(Yang *et al.*, 2012)
Barley				
Sprouting detection	1430	PCA*; MNF	97%**	(Arngren *et al.*, 2011)
Sprouting detection	1920–1940	PCA*; PLS-DA	0.68	(McGoverin *et al.*, 2011)
Sorghum				
Sprouting detection	1920–1940	PCA*; PLS-DA	0.53	(McGoverin *et al.*, 2011)
General				
Grain cleanness	465–475; 522– 532; 676–705; 849–858; 906–945	PLS-DA	-	(Wallays *et al.*, 2009)

*** PCA used for exploratory statistical analysis only**

**** Classification accuracy (R^2 not available)**

McGoverin and Manley (2012) developed PCA classification maps of kernels where regions of white maize kernels were characterised as germ, vitreous and floury endosperm, expressed as a percentage of the whole image. Manual dissection of the kernels confirmed the result of the NIR hyperspectral image analysis, with an R^2 of 0.61, 0.59 and 0.11 for vitreous endosperm, floury endosperm and germ, respectively. The incongruence between the manual method and image analysis is likely due to the surface–bias of reflectance spectroscopy, where results are based on shallow measurements only, and not proportions deeper within the kernel.

Wheat kernel hardness has been determined, classifying individual kernels as either vitreous, partial vitreous, and/or starchy (Gorretta *et al.*, 2006). Three wavelengths related to the wheat kernel components were found to be significant, namely protein at 910 nm, starch at 990 nm and amino acids at 1030 nm. A PLS factorial discriminant analysis (PLS-FDA) model with a high classification accuracy of 94% was developed.

The use of PLS regression and PCR were compared for the prediction of hardness values of bulk samples of wheat (Mahesh *et al.*, 2015). The comparison of the two methods was based on the estimated mean square errors of prediction (MSEP), standard error of cross-validation (SECV), and R^2. PLS regression method performed superiorly to the PCR method regarding all of these parameters. It was found that the wavelength regions of 1180–1220; 1320–1400; and 1460–1490 nm were critical for predicting hardness using the PLSR method, while 960–1030 and 1670–1700 nm were identified as important from the loading values of PC1 in the PCR model. Furthermore, absorptions in the wavelength regions of 960–1060; 1330–1480; and at 1680 nm are likely associated to protein, starch, and hardness of samples, respectively.

Manley *et al.* (2011) studied the effect of different hardness on the diffusion of conditioning water into wheat kernels. Wheat samples of known hardness were used, and both deionised water and deuterium oxide (D_2O) were experimented with as the conditioning water. It was found that there was a low sensitivity of detection for D_2O, and changes were not clear. The deionised water samples performed better with regards to detection by NIR hyperspectral imaging, and changes over time were more easily detected. PCA was calculated up to PC4 and 8 wavelengths related to the ionised water samples were found to be significant. Interestingly, four of the wavelengths (1886, 1899, 1936 and 1984 nm) were found to indicate whether the water was bound or free, and indicated the exact manner in which the bound water had formed its hydrogen bonds, displaying how powerful NIR hyperspectral imaging can be.

Chemical Composition

There is an increasing need in agricultural science for information on specific traits in cereal samples, driving the need to find high-throughput and non-destructive methods of analysis. As cereal kernels are heterogeneous in nature, conventional spectroscopic determination of chemical composition may be difficult without the use of hyperspectral imaging. Spatially resolved spectra

allow researchers to study the distribution of the various chemical components identified. Not only is knowledge of a kernel's chemical composition important for quality classification for processing purposes, but also for breeding where limited material is available. Single kernels thus need to be analysed non-destructively to enable propagation of next generation material.

Lin *et al.* (2006) determined the moisture content of rice using MLR, PLS regression and ANN. The PLS regression model had very similar prediction ability (R^2 = 0.94; RMSEP = 0.48%) to the MLR approach (R^2 = 0.95; RMSEP = 0.44%). However, it required more input wavelengths. The ANN model had performance similar to the others (R^2 = 0.95; RMSEP = 0.44%), demonstrating that it can provide comparable detection capability for rice moisture.

Moisture and oil concentration in single maize kernels have been determined using PLS regression and PCR (Cogdill *et al.*, 2004). Several pre-processing techniques were experimented with to optimise model performance. The moisture calibration results were better than that of the oil calibration. PCR performed slightly worse than PLS regression, but the stability of the PCR prediction was far superior to that of PLS. For the moisture calibration, PLS regression calculated on the raw absorbance spectra with 11 PLS factors achieved an R^2 of 0.871 and RMSEP of 0.87%, while the PCR model had an R^2 of 0.856 and very low RMSEP of 0.07%. The inaccurate results obtained from the oil predictions were declared hardly useful, where calibration and reference chemistry errors accounted for nearly 50% and 40% of the large overall error, respectively. However, upon inspection of the results, the calibration appeared to have some potential, and it was suggested that a larger or more diverse training set should have been used so that error arising from the conventional testing of the reference samples can be overcome by volume.

Similarly, Weinstock *et al.* (2006) predicted oil and oleic acid concentrations in single maize kernels using PLS regression models. Oil content was predicted with an RMSEP of 0.74% and R^2 of 0.68, and oleic acid content with an RMSEP of 13.7% and R^2 of 0.65. Both of the models for predicting the respective concentrations did not perform very well, but in comparison to the study by Cogdill *et al.* (2004), these results were a vast improvement. By demonstrating that the technique could be developed further for predicting concentrations of minor chemical components, one can recognise the potential to be applied to further phenotypic traits, thus replacing low-throughput, laborious analytical methods currently used.

The use of PLS regression and PCR were compared for the prediction of protein content of bulk samples of wheat (Mahesh *et al.*, 2015). The comparison of the two methods was based on MSEP, SECV, and R^2. PLS regression method performed superiorly to the PCR method regarding all of these parameters. The wavelength regions of 1180–1200 and 1460–1500 nm were found to be important features in the PLSR method, while 960–1030 and 1670–1700 nm were vital for the first PC loading values in the PCR method. The NIR absorptions of protein contents of samples were seen at the wavelengths of 1470–1500 nm.

Variety Identification

Varieties of cereal grains tend to have very similar appearances, and manually differentiating varieties based on these subtle visual differences may be challenging, inconsistent and labour intensive. However, the properties of the grain varieties may vary dramatically regarding their chemical composition and physical characteristics, and consequently also their end-product quality. Alternatively, it is possible to use the physiochemical differences as a means of differentiating between varieties. Object-wise classification maps can be used to visualise the variety of a kernel using mathematical models, where classified objects with similar physiochemical composition are shown as the same colour in the image (Kucheryavskiy, 2013). This has great potential for online or real-time industry quality control and grading, as a grader can make use of an automated system to classify different varieties non-destructively.

Visual differentiation can be replaced with an NIR hyperspectral imaging system by developing classification models to differentiate wheat varieties. A study by Mahesh *et al.* (2008) classified eight wheat varieties using three multivariate approaches. R^2 values of 0.94 to 1.00 and 0.86 to 1.00 were obtained from LDA and QDA models, respectively. The R^2 values of the ANN models ranged 0.80 to 1.00, with less than 10% error.

In a similar study, wheat classes were differentiated using three chemometric techniques, LDC, QDA and BPNN, with classification accuracies of 99, 98 and 92%, respectively (Choudhary *et al.*, 2009). Wavelet texture analysis was applied to investigate whether textural features of the hyperspectral images have similarities within a wheat class but differ between classes. These features were probably related to scattering differences on or just below the surface. Thus, this approach focuses on the physical differences between varieties, as opposed to chemical composition. For each waveband, 3000 wavelet features were obtained by wavelet analysis, which needed to be reduced dramatically. The best fitting LDC, QDA and BPNN models required the top 90, 90 and 70 visual texture features, respectively. The important wavelengths within the top 100 features were 960, 970 and 980 nm related to moisture content; 1160, 1680 and 1700 nm related to carbohydrate content; and 1450 nm, related to protein content. Wavelet texture analysis of NIR hyperspectral images was shown to be effective for wheat class discrimination (Choudhary *et al.*, 2009).

Discrimination of three rice varieties was modelled using PCA and various BPNN models (Wang *et al.*, 2014a). First, a simple PCA unsupervised classification approach was used. The varieties were easily distinguished from one another in the principal component space and a relatively high classification accuracy of 89% was achieved. Supervised classification has a potential to offer better predictability, and was carried out using BPNN in three different ways, based on optimised spectral data, on image information, and on both spectral data and image information. For the spectral based model, seven optimal wavelengths, namely 5 visible (419, 452, 593, 613 and 743) and 2 NIR (784 and 966 nm), were selected for the basis of discrimination and the resulting

classification accuracy was 90%. The lower classification accuracy (88%) of the image information based model demonstrated that the spectral information contained more feature information. The model based on the fusion of spectral data and image information obtained the highest classification accuracy of 94%. This demonstrated that supervised classification with BPNN is superior to PCA, especially when based on both the spectral and image information, and thus both chemical and physical kernel properties. Recently, the same researchers reported on the differentiation of maize varieties (Wang *et al.*, 2015). As in previous variety identification studies (Choudhary *et al.*, 2009), wavelet analysis was utilised to provide texture features which was combined with spectral data to improve classification accuracy. Three least squares SVM models were calculated, based on only spectral, only textural, or combined spectral and textural data. The combined model performed best with a classification accuracy of 89%, followed by the spectral model (87%) and textural model (85%). However, the results for rice were found to be more precise, which may be due to differences in the cereals or to the different data analyses. It is known that spectral information can present some significant internal chemical attributes of maize kernels, which are closely related to the varieties. An example is waxy maize and sweet maize, whose main constituents are amylopectin and sugar, respectively. Differences in textural features may also play a role in classifying varieties, however the textural models developed based on image data were described as unsatisfactory, and thus it is unclear if combining spectral and textural data was truly beneficial.

Sprouting Detection

Sprouting, or pre-germination, is a prevalent defect occurring when germination begins due to exposure to water, often due to rain prior to harvest (McGoverin *et al.*, 2011). Sprouting may leave the kernel more susceptible to microbial contamination and insect infestation. Enzymatic activity increases, causing starch to be converted to sugars, changes in protein composition, and a loss in dry matter content (Singh *et al.*, 2009b). An important enzyme related to sprouting in wheat is α-amylase, where the increased concentrations negatively affect the baking quality of wheat flour used in the baking of bread (Van Der Maarel *et al.*, 2002). This is further affected by the reduced starch content. Grain weight and yield are also negatively affected. The standard test for sprouting is a Falling Number test. Flour and water are mixed to test the time required for a plunger to sink through the mixture. A fast falling number (300 seconds or less) indicates a high degree of sprouting. Another method of detecting sprouted kernels is to induce germination in a sample, as sprouted kernels are unable to germinate a second time. Alternatively, the kernels may be visually inspected, but early detection is not possible. All methods are very time-consuming.

NIR hyperspectral imaging was found to be superior to methods commonly used to discriminate between healthy wheat kernels, artificially sprouted kernels, and midge-damaged kernels, which are often mistaken for sprouted kernels during visual inspections (Singh *et al.*, 2009b). The wavelengths 1102 and 1132 nm were selected as significant wavelengths, associated

with starch. In addition, 1305 nm was also found to be significant. Models based on LDA, QDA, and Mahalanobis discriminant classifiers were developed. The Mahalanobis discriminant classifier is a simplified form of LDA with assumptions that all known classes have equal classification probabilities. All three of the discriminant analyses detected 100% of the artificially sprouted kernels. However, these artificial samples were severely sprouted, and thus were easily detected. This may not necessarily be the case for naturally sprouted wheat. The Mahalanobis classifier gave the highest and most consistent classification of the midge-damaged samples (average classification accuracy = 98%) compared to LDA (93%) and QDA (92%). The LDA (95%) and QDA (98%) better classified healthy kernels, compared to the Mahalanobis classifier (75%). The classification accuracy of crease-up orientated samples was slightly higher than crease-down samples, because these defects occur non-uniformly, and are most prevalent on the crease side.

In brewing, germination of barley is required in malting, one of the initial processing steps. The major defect of pre-germinated barley is its predisposed inability to germinate again, and must thus be used for a lower grade such as animal feed. Using NIR hyperspectral imaging, one is able to identify pre-germinated barley at an early stage of approximately 12 h after of pre-germination (Arngren *et al.*, 2011). The MNF model developed, classified kernels as either normal, late germinating or limited/no germinating. A high classification accuracy (97%) was achieved on a bulk kernel level. The model provided the assigned class and probability for each kernel, which aids in separating kernel classes and attaining homogeneous germination profiles.

McGoverin *et al.* (2011) was able to differentiate between viable and sprouted barley, wheat and sorghum using PLS-DA classification. PCA was applied and distinct clustering was not observed (Fig. 2.4). However, a gradual change of scores with incubation time was observed in the score images. Positive PC5 score values were largely associated with viable barley kernels, and the negative values with sprouted kernels. The image dataset was divided accordingly into viable and sprouted classes. The R^2 values for the PLS-DA models were not very high, where average R^2 values were 0.68, 0.67 and 0.53 for barley, wheat and sorghum, respectively. The classification resulted in large numbers of false positives, false negatives and unclassified kernels. Interpretation of the PCA loading line plots found an O-H stretching and deformation band (1920–1940 nm) prominent in all groups, as it is associated with water binding, a major contributor to the viable/sprouted differentiation.

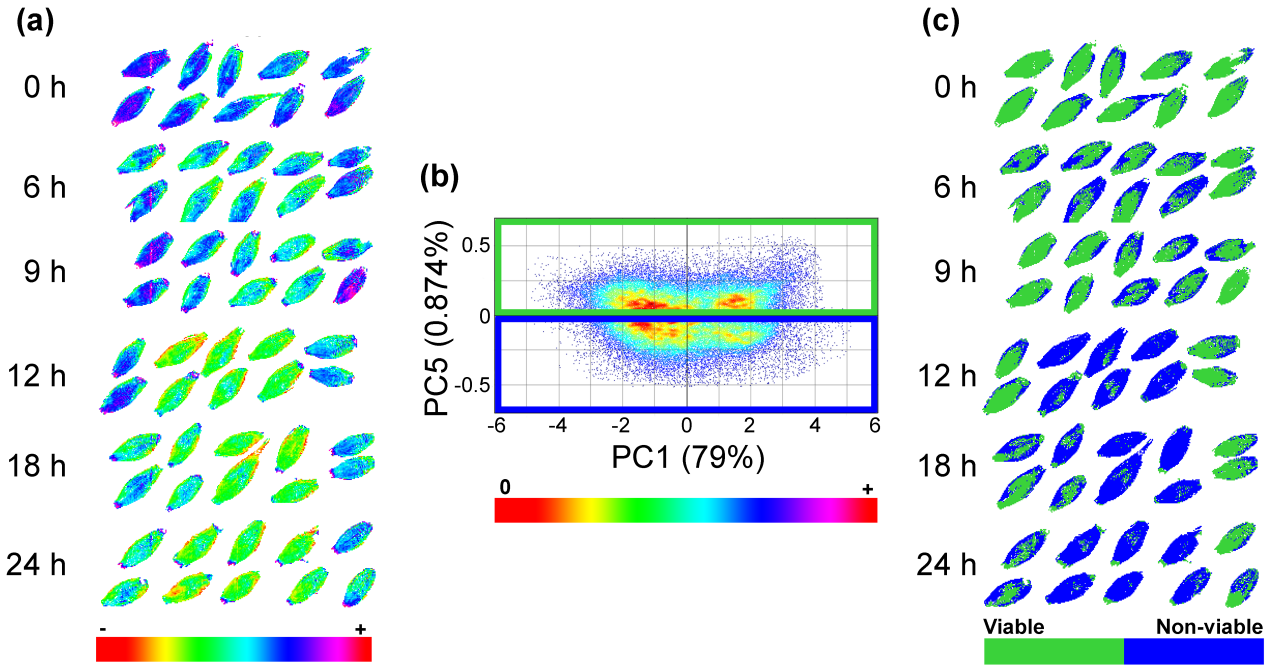

Figure 2.4 (a) PC5 score image for the barley cultivar Erica image dataset where dark blue indicates viable kernels and green non- viable kernels. (b) The corresponding PC5 vs. PC1 score plot showing density and clustering of pixels. Positive PC5 scores values (green box) were identified and assigned to the viable class and negative PC5 scores values (blue box) to the non-viable class. Brushing between score plot and score image was used. (c) Classification image showing viable and non-viable regions. [Copyright 2011 Springer. Reproduced with permission from McGoverin et al. (2011)].

Physical Quality Classification

One of the earliest applications of NIR hyperspectral imaging in cereal sciences was the classification of wheat kernel colour by Archibald *et al.* (1998). Eleven wavelengths were selected as significant in differentiating between hard red spring and hard winter white varieties. However, the PCA model performed quite poorly, and it was found that differentiation was due to only one spectral feature. The study demonstrated distinct advantages over bulk or single-kernel spectroscopy, but left room for future improvement with regard to experimental technique and data analysis. Further multivariate techniques, such as PLS regression or PLS-DA should have been applied to utilise multiple spectral features in order to classify or quantify the colour of each kernel.

A parameter important in grading is often the amount of foreign material. By calculating the proportion of foreign material pixels in an NIR hyperspectral image of grain, a measure for the weight percentage foreign material in the sample was obtained. Wallays *et al.* (2009) developed a PLS-DA model using five visible and NIR spectral regions (465–475 nm, 522–532 nm, 676–705 nm, 849–858 nm and 906–945 nm) to allow classification of each individual pixel with a relative prediction error of 0.26.

Safety

Fungal Contamination Detection

Despite the majority of crops being treated with fungicides, disease problems continue to develop in many areas, affecting both the yield and quality of the grain crop. *Fusarium* spp. infections are the most economically significant diseases, causing ear rots in maize, and head blight in wheat and barley (Munkvold and Desjardins, 1997). The mycotoxins produced, fumonisins in the case of *Fusarium* spp., are harmful to humans and animals, and are thus often controlled by legal limit values.

A rapid, non-destructive method for the early detection of potentially toxigenic fungal species in maize has been developed, which showed great promise for on-line safety and quality control (Del Fiore *et al.*, 2010). Various *Aspergillus* and *Fusarium* strains were investigated, and differentiation of kernels infected with these fungi from sound grains, as well as between the two fungal species, was achieved. The most notable result was the early discrimination of both *Aspergillus flavus* and *Aspergillus niger* infected kernels from sound kernels, just 48 h after inoculation. A specific spectral profile, or spectral signature, was produced for each species, with four wavelengths (410, 470, 535 and 945 nm) found to be significant. It was suggested that the spectral differences observed at wavelengths below 700 nm were caused by the spectral properties of the fungi. However, above 700 nm and into the NIR range, the spectral differences were determined by scattering arising due to changes on the surface of the kernels because of the onset of fungal infection.

PCA and PLS-DA models were used by Williams *et al.* (2012) who was able to detect *Fusarium* infection in maize kernels within a quicker time frame of 20 h. In the PCA models, three distinct clusters were apparent in the score plot (Fig. 2.5). There was a distinct decrease in score values from the control to infected kernels. This is illustrated by the change in colour from red (control) to blue (infected). A classification plot was generated by assigning each of the three clusters to a class in the score plot, which, when projected onto the score image, formed a classification image. Green was associated with the control samples, black was associated with early time intervals (<20 h), and red was associated with the later time intervals (>20 h). These changes can be described as an infection gradient, giving an indication to the extent of infection. The PLS regression model was calculated with time as the y-variable. The coefficients of determination for the model were not very high (R^2 = 0.71 and 0.63), and had RMSEPs of 7.0 h and 8.2 h, respectively. The wavelengths found to be important to the differentiation of infected kernels were 1405 nm, 1660–1668 nm, 1900 nm, and 2136 nm. It was deduced that 1405 nm was associated with starch, 1668 nm with the plastic Petri-dish, and 2136 nm with amino acid, and thus protein. These features were associated with changes in the kernel as a result of kernel constituents being depleted during fungal proliferation and were not associated with the fungal growth itself.

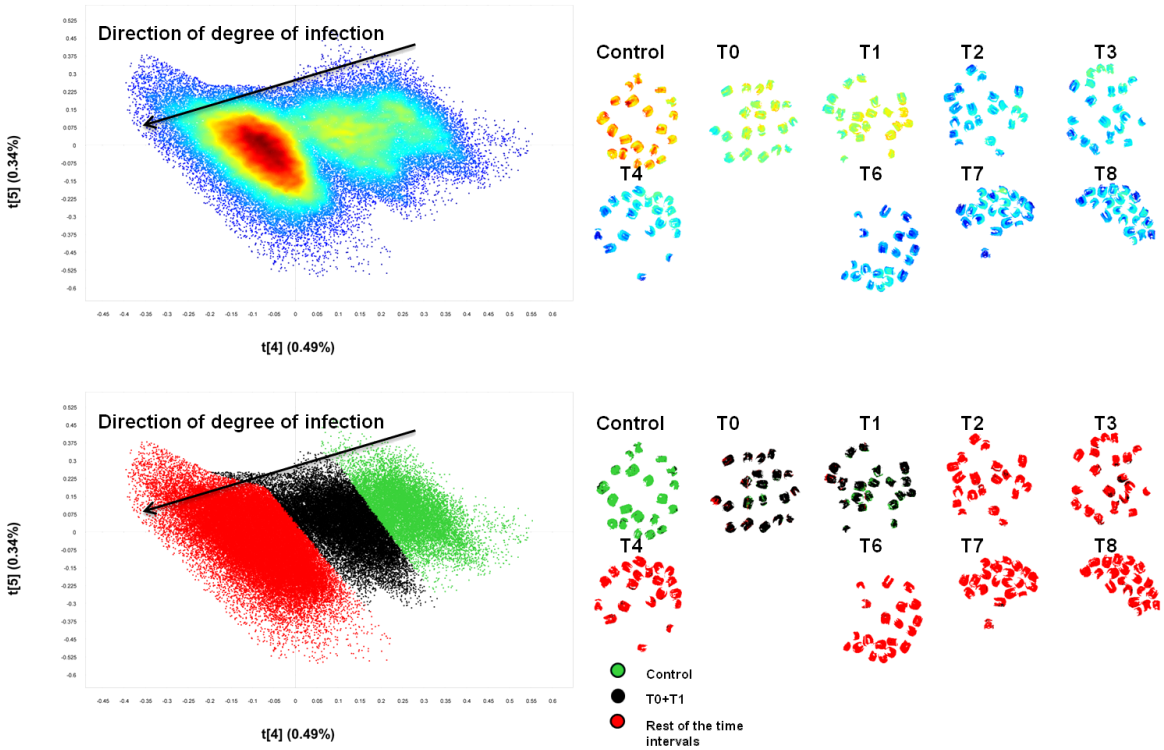

Figure 2.5 (a) PCA score plot of PC4 vs. PC5 (0.49% and 0.34% SS) showing three clusters and (b) corresponding score image of PC4 showing decrease in score values from left to right, where warm colours (yellow to red) correspond with positive score values and cold colours (blue) correspond to negative score values; (c) classification plot of PC4 vs. PC5 with classes and degree of infection in the direction of PC4 and (d) classes projected onto score image showing the control (green), T0&T1 (black) and the remaining time intervals up to 90 h post inoculation (red). [Copyright 2012 Elsevier. Reproduced with permission from Williams et al. (2012)].

Fusarium infection in wheat has also been detected, however it was found that infection was not detectable immediately after inoculation, and the onset of symptoms was required for detection (Bauriegel *et al.*, 2011). SAM classification was applied to identify four wavelength ranges that allow diseased and healthy ear tissues to be differentiated with a classification accuracy of 87%. The spectral differences appearing in the visible and NIR regions were attributed to disease development. Most significant were changes in the content of carotenoids (500–533 nm), the content of chlorophylls (560–675 nm and 682–733nm), and tissue water content (927–931 nm). In a similar study, stained and sound wheat kernels have been differentiated using NIR spectroscopy and visible hyperspectral imaging (Berman *et al.*, 2007). Staining was either grey/black due to visible fungal growth, or pink due to *Fusarium* infection. Black-point staining, due to a stress-induced phenolic oxidation reaction, was also included. All three stained categories were differentiated from one another and from sound wheat kernels. Image acquisition was conducted using three wavelength ranges, which included NIR and visible regions, namely the full wavelength range (420–2500 nm), as well as two reduced ranges (420–1000 nm and 420–700 nm). Results for all three

were very similar, suggesting that a smallest range may be used without forfeiting high classification accuracy (95%), increasing the ease of adapting the technique to industry use.

NIR hyperspectral imaging has also been used to detect the presence of aflatoxin B_1 on wheat kernels, where it was reported that the concentrations detected were as low as 10 ppb (Wang *et al.*, 2014b). This study was different from previously mentioned studies as the presence of mycotoxin was tested as an indicator for fungal contamination, whereas the others observed changes in the kernels as a result of infection. The FDA models developed differentiated the infected and uninfected kernels with a classification accuracy of 94%. Significant wavelengths included 2344 nm, associated with the coumarin ring in aflatoxin B_1, and 2057 nm, associated with the lactone moiety of aflatoxin B_1. An issue with this study is that the detection of concentrations below 1000 ppm using this technique is often contested. It could be hypothesised that kernels with levels of 10 ppb and higher exhibited additional unique features which were the basis on which the model was calibrated. Although it may not be correct to claim that aflatoxin B_1 is predicted in this study, the calibration could still be useful.

Another toxigenic fungus affecting several crops is *Claviceps purpurea*. It produces sclerotia or ergot bodies, small resting stages that can end up in grain products and cause severe human health issues. Vermeulen *et al.* (2012) detected and quantified ergot contamination in wheat. Pixels were detected as either ergot or wheat using PLS-DA models, and the ergot concentration was calculated as a percentage. The results showed a very high correlation between the predicted values obtained using models based on both PLS-DA and SVM, and the reference values (R^2 > 0.99). The study demonstrated that NIR hyperspectral imaging is sensitive enough to detect 500 mg/kg (the maximum ergot level for human food), in line with European legislation.

Rice Blast (*Pyricularia*) is a devastating fungal disease affecting rice. A method to detect the early stages of the disease was developed by Yang *et al.* (2012). To simplify analysis and reduce costs of instrumentation for the prospective practical applications, a minimal number of wavelengths is desirable. Five wavelengths (1188, 1339, 1377, 1432 and 1614 nm) were found to be significant and were used to develop an LDA model to differentiate infected kernels with a classification accuracy of 92%. This result is encouraging, indicating that information from a few wavelengths to identify the rice blast disease, and perhaps other cereal properties, is feasible. This method could potentially allow farmers to choose to treat only infected crops with fungicide to control the onset of the disease. Thus, entire crops are not sprayed unless there is a confirmed infection, reducing financial cost and resulting in lower chemical residues.

Parasitic Contamination Detection

It is estimated that 10 to 30% of worldwide grain production is lost every year due to insect and rodent damage (Singh *et al.*, 2009a). Much of this damage occurs during storage, and will lead to a loss of weight, nutrients, and germination ability, and increased susceptibility to contamination by

fungi. In turn, much of the market value of the grain is lost. Furthermore, insect debris is likely to be incorporated into the processed products from the contaminated grain, where legal limits stipulate a maximum level of insect fragments as part of foreign material upper limits. Internal feeding insects develop inside the kernel, after which the adult emerges via an exit hole, producing an insect-damaged kernel (Singh *et al.*, 2010). These insects generally cause the most damage, and are difficult to detect. Secondary feeding insects will subsequently feed on the broken and damaged kernels. The metabolic activity of the insects produces moisture and heat, which provides an optimal environment for fungal contamination to develop.

Singh *et al.* (2009a) detected wheat kernels contaminated by *Sitophilus oryzae*, *Rhyzopertha dominica*, *Cryptolestes ferrugineus*, and *Tribolium castaneum* using the wavelength range 1000–1600 nm. Classification algorithms were developed with LDA, QDA, and Mahalanobis discriminant classifiers. However, the Mahalanobis classifier performed very poorly and inconsistently, and the results were thus not included in the reporting of the study. Initially, binary classification models were developed, and gave very high classification accuracies. However, naturally infested wheat is likely to be infested with more than one insect species. Thus, the sound kernels were grouped into one class and kernels damaged by all four insect species into another. Several two-way classification models were developed, but the best results arose from combining with the histogram features and the statistical features of the first PC wavelengths, 1101.69 and 1305.05 nm, both of which are related to absorption by starch molecules. The LDA model gave the highest and most consistent classification of the sound kernels (average classification accuracy = 98%) compared to QDA (93%). The QDA (100%) better classified the damaged kernels, compared to the LDA (94%). The classification accuracy of crease-up orientated samples was slightly higher than crease-down samples. These authors followed up with a similar study (Singh *et al.*, 2010), using the same insects as contaminants, but made use of a shorter wavelength range (700–1100 nm) instrument, as well as a second colour imaging system. Again, LDA, QDA, and Mahalanobis discriminant analyses were conducted. In the NIR hyperspectral image data analyses, the Mahalanobis discriminant classifier gave the highest classification accuracy in detecting the insect-damaged kernels (98%), above LDA (95%) and QDA (89%), while QDA gave the highest classification accuracy in classifying sound kernels (86%). For the colour images, morphological, colour, and textural image information was extracted. However, only the colour features seemed to possess a high discrimination capability, and were thus used only. LDA gave the highest classification accuracy for the discrimination of sound kernels (80%) and insect-damaged kernels (94%). However, for all models, it was found that there were a large number of false positives occurring, where healthy kernels were deemed damaged.

CONCLUSION

The current grading methods used by the South African maize industry could benefit from the use of complementary analytical techniques. Spectral imaging, specifically NIR hyperspectral imaging, has been identified as a potential method for improved maize grading. NIR hyperspectral imaging integrates spectroscopy and computer imaging, and spectral and spatial information is acquired simultaneously. The richness of the acquired information allows for various multivariate data analyses to be applied. Most works deal with multivariate techniques to reduce the vast size of hyperspectral image datasets, with the most cited being based on PCA, PLS-DA, LDA and PLS regression. A wide range of properties may be studied to give a rapid, reliable and accurate description of grain's internal and external quality and safety status. Currently, hyperspectral instruments are possibly better suited to laboratory work, but most studies are working toward providing the industry with important practical solutions. This review shows promise for the grading of maize using spectral imaging, as previous research has established the capability of spectral imaging for real-time, on-line industry classification and sorting of cereals. However, further reduction in instrumentation costs, and improvements in computation capacity, expeditiousness, sensitivity, stability, scanning rate, and portability of instruments are required.

REFERENCES

Amigo, J. M., Martí, I. & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality. In: *Data Handling in Science and Technology*. Pp. 343-370. Amsterdam, The Netherlands: Elsevier Science.

Archibald, D. D., Thai, C. N. & Dowell, F. E. (1998). Development of short-wavelength near-infrared spectral imaging for grain color classification. In: *Photonics East (ISAM, VVDC, IEMB)*. Pp. 189-198. International Society for Optics and Photonics.

Arngren, M., Hansen, P. W., Eriksen, B., Larsen, J. & Larsen, R. (2011). Analysis of pregerminated barley using hyperspectral image analysis. *Journal of Agricultural and Food Chemistry*, **59**, 11385-11394.

Baeten, V., Pierna, J. A. F. & Dardenne, P. (2007). Hyperspectral Imaging Techniques: an Attractive Solution for the Analysis of Biological and Agricultural Materials. In: *Techniques and applications of hyperspectral image analysis* (edited by H. Grahn & P. Geladi). Pp. 289-312. Chichester, UK: John Wiley & Sons.

Barnes, R. J., Dhanoa, M. S. & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, **43**, 772-777.

Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U. & Herppich, W. B. (2011). Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture*, **75**, 304-312.

Berman, M., Connor, P. M., Whitbourn, L. B., Coward, D. A., Osborne, B. G. & Southan, M. D. (2007). Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **15**, 351-358.

Blandino, M., Mancini, M. C., Peila, A., Rolle, L., Vanara, F. & Reyneri, A. (2010). Determination of maize kernel hardness: comparison of different laboratory tests to predict dry‐milling performance. *Journal of the Science of Food and Agriculture*, **90**, 1870-1878.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, **131**, 1152-1160.

Chevallier, S., Bertrand, D., Kohler, A. & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, **20**, 221.

Choudhary, R., Mahesh, S., Paliwal, J. & Jayas, D. S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering*, **102**, 115-127.

Cogdill, R. P., Hurburgh, C. R., Rippke, G. R., Bajic, S. J., Jones, R. W., McClelland, J. F., Jensen, T. C. & Liu, J. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE*, **47**, 311.

Cowe, I. A. & McNicol, J. W. (1985). The use of principal components in the analysis of near-infrared spectra. *Applied Spectroscopy*, **39**, 257-266.

Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V. & Pierna, J. A. F. (2013). Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. *Applied Spectroscopy Reviews*, **48**, 142-159.

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64-71.

Department of Agriculture (2016). Regulations relating to the grading, packing and marking of bread wheat intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Elmasry, G., Kamruzzaman, M., Sun, D. W. & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical Reviews in Food Science and Nutrition*, **52**, 999-1023.

Elmasry, G. & Sun, D. W. (2010). Principles of hyperspectral imaging technology. In: *Hyperspectral imaging for food quality analysis and control* (edited by D.W Sun.). Pp. 3-44. London, UK: Elsevier.

Elmasry, G., Sun, D. W. & Allen, P. (2011). Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International*, **44**, 2624-2633.

Esbensen, K. & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, **7**, 67-86.

Esbensen, K. H., Guyot, D., Westad, F. & Houmoller, L. P. (2002). Principal Component Analysis (PCA) - Introduction. In: *Multivariate data analysis - In practice: An introduction to multivariate data analysis and experimental design*. Pp. 19-74. Camo ASA.

Fandohan, P., Hell, K., Marasas, W. F. O. & Wingfield, M. J. (2004). Infection of maize by Fusarium species and contamination with fumonisin in Africa. *African Journal of Biotechnology*, **2**, 570-579.

Feng, Y. Z. & Sun, D. W. (2012). Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in Food Science and Nutrition*, **52**, 1039-1058.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179-188.

Fox, G. & Manley, M. (2009). Hardness methods for testing maize kernels. *Journal of Agricultural and Food Chemistry*, **57**, 5647-5657.

Giroux, M. J. & Morris, C. F. (1997). A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. *Theoretical and Applied Genetics*, **95**, 857-864.

Goetz, A. F. H. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment*, **113**, 5-16.

Goetz, A. F. H., Vane, G., Solomon, J. E. & Rock, B. N. (1985). Imaging spectrometry for earth remote sensing. *Science*, **228**, 1147-1153.

Gorretta, N., Roger, J. M., Aubert, M., Bellon-Maurel, V., Campan, F. & Roumet, P. (2006). Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **14**, 231-239.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G. & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Hege, E. K., O'Connell, D., Johnson, W., Basty, S. & Dereniak, E. L. (2004). Hyperspectral imaging for astronomy and space surviellance. In: *Optical Science and Technology, SPIE's 48th Annual Meeting*. Pp. 380-391. International Society for Optics and Photonics.

Jay, S. C., Lawrence, R. L., Repasky, K. S. & Rew, L. J. (2010). Detection of leafy spurge using hyper-spectral-spatial-temporal imagery. In: *Geoscience and Remote Sensing Symposium* (IGARSS). Pp. 4374-4376. IEEE.

Keith, C. J., Repasky, K. S., Lawrence, R. L., Jay, S. C. & Carlsten, J. L. (2009). Monitoring effects of a controlled subsurface carbon dioxide release on vegetation using a hyperspectral imager. *International Journal of Greenhouse Gas Control*, **3**, 626-632.

Koohafkan, P. & Stewart, B. A. (2008). Introduction. In: *Water and Cereals in Drylands*. Pp. 1-5. London, U.K.: Earthscan.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126-135.

Li-Chan, E. (2011). Introduction to vibrational spectroscopy in Food Science. In: *Applications of vibrational spectroscopy in Food Science* (edited by Li-Chan, E., Chalmers, J. & Griffiths, P.). Pp. 3-30. Chichester, UK: John Wiley & Sons.

Lin, L. H., Lu, F. M. & Chang, Y. C. (2006). Development of a near-infrared imaging system for determination of rice moisture. *Cereal Chemistry*, **83**, 498-504.

Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O. L. & Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, **5**, 1121-1142.

Mahesh, S., Jayas, D. S., Paliwal, J. & White, N. D. G. (2015). Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat. *Food and Bioprocess Technology*, **8**, 31-40.

Mahesh, S., Manickavasagan, A., Jayas, D. S., Paliwal, J. & White, N. D. G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101**, 50-57.

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews*, **43**, 8200-8214.

Manley, M., Du Toit, G. & Geladi, P. (2011). Tracking diffusion of conditioning water in single wheat kernels of different hardnesses by near infrared hyperspectral imaging. *Analytica Chimica Acta*, **686**, 64-75.

Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (Zea maize L.) kernels. *Journal of Agricultural and Food Chemistry*, **57**, 8761-8769.

Masango, M., Flett, B., Ellis, C. & Botha, C. (2014). Stenocarpella maydis and its toxic metabolites: a South African perspective on diplodiosis. *World Mycotoxin Journal*, **8**, 341-350.

McGoverin, C. & Manley, M. (2012). Classification of maize kernel hardness using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **20**, 529.

McGoverin, C. M., Engelbrecht, P., Geladi, P. & Manley, M. (2011). Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. *Analytical and Bioanalytical Chemistry*, **401**, 2283-2289.

Mdangi, M., Mulungu, L. S., Massawe, A. W., Eiseb, S. J., Tutjavi, V., Kirsten, F., Mahlaba, T., Malebane, P., von Maltitz, E. & Monadjem, A. (2013). Assessment of rodent damage to stored maize (Zea mays L.) on smallholder farms in Tanzania. *International Journal of Pest Management*, **59**, 55-62.

Munkvold, G. P. & Desjardins, A. E. (1997). Fumonisins in maize: Can we reduce their occurrence? *Plant Disease*, **81**, 556-565.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). Data compression by PCR and PLS. In: *A user-friendly guide to multivariate calibration and classification*. Pp. 27-38. Chichester, U.K.: NIR Publications.

Nansen, C., Kolomiets, M. & Gao, X. (2008). Considerations regarding the use of hyperspectral imaging data in classifications of food products, exemplified by analysis of maize kernels. *Journal of Agricultural and Food Chemistry*, **56**, 2933-2938.

Nansen, C., Zhao, G., Dakin, N., Zhao, C. & Turner, S. R. (2015). Using hyperspectral imaging to determine germination of native Australian plant seeds. *Journal of Photochemistry and Photobiology*, **145**, 19-24.

Ominski, K. H., Marquardt, R. R., Sinha, R. N. & Abramson, D. (1994). Mycotoxins in grains - Compounds other than aflatoxin. In: *Ecological aspects of growth and mycotoxin production by storage fungi*. Pp. 287-305. St. Paul, USA: Eagan Press.

Patterson, M. C. L. & Brescia, A. (2010). Operation of small sensor payloads on tactical sized unmanned air vehicles. *Aeronautical Journal*, **114**, 427-436.

Polerecky, L., Bissett, A., Al-Najjar, M., Faerber, P., Osmers, H., Suci, P. A., Stoodley, P. & de Beer, D. (2009). Modular spectral imaging system for discrimination of pigments in cells and microbial communities. *Applied and Environmental Microbiology*, **75**, 758-771.

Ravn, C., Skibsted, E. & Bro, R. (2008). Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches. *Journal of Pharmaceutical and Biomedical Analysis*, **48**, 554-561.

Resmini, R. G., Kappus, M. E., Aldrich, W. S., Harsanyi, J. C. & Anderson, M. (1997). Mineral mapping with hyperspectral digital imagery collection experiment (HYDICE) sensor data at Cuprite, Nevada, USA. *International Journal of Remote Sensing*, **18**, 1553-1570.

Serna-Saldivar, S. O. (2010). Physical Properties, Grading, and Speciality Grains. In: *Cereal grains: properties, processing, and nutritional attributes*. Pp. 43-80. Boca Raton, USA: CRC Press.

Singh, C. B., Jayas, D. S., Paliwal, J. & White, N. D. G. (2009a). Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. *Journal of Stored Products Research*, **45**, 151-158.

Singh, C. B., Jayas, D. S., Paliwal, J. & White, N. D. G. (2009b). Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry*, **86**, 256-260.

Singh, C. B., Jayas, D. S., Paliwal, J. & White, N. D. G. (2010). Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Computers and Electronics in Agriculture*, **73**, 118-125.

Spangler, L. H., Dobeck, L. M., Repasky, K. S., Nehrir, A. R., Humphries, S. D., Barr, J. L., Keith, C. J., Shaw, J. A., Rouse, J. H. & Cunningham, A. B. (2010). A shallow subsurface controlled release facility in Bozeman, Montana, USA, for testing near surface CO2 detection techniques and transport models. *Environmental Earth Sciences*, **60**, 227-239.

Sydenham, E. W., Thiel, P. G., Marasas, W. F., Shephard, G. S., Van Schalkwyk, D. J. & Koch, K. R. (1990). Natural occurrence of some Fusarium mycotoxins in corn from low and high esophageal cancer prevalence areas of the Transkei, Southern Africa. *Journal of Agricultural and Food Chemistry*, **38**, 1900-1903.

Symons, S. J. & Shahin, M. A. (2008). Quality Evaluation of Corn/Maize. In: *Computer Vision Technology for Food Quality Evaluation* (edited by D.W. Sun). Pp. 401-424. Canada: Elsevier.

Thenkabail, P. S. (2012). Foreword. In: *Hyperspectral remote sensing of vegetation* (edited by P. S. Thenkabail, J. G. Lyon & A. Huete). Pp. xi-xii. Boca Raton, USA: CRC Press.

Van Der Maarel, M. J. E. C., Van Der Veen, B., Uitdehaag, J. C. M., Leemhuis, H. & Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the α-amylase family. *Journal of Biotechnology*, **94**, 137-155.

Vermeulen, P., Pierna, J. A. F., Van Egmond, H. P., Dardenne, P. & Baeten, V. (2012). Online detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging. *Food Additives & Contaminants*, **29**, 232-240.

Vidal, M. & Amigo, J. M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, **117**, 138-148.

Wallays, C., Missotten, B., De Baerdemaeker, J. & Saeys, W. (2009). Hyperspectral waveband selection for on-line measurement of grain cleanness. *Biosystems Engineering*, **104**, 1-7.

Wang, L., Liu, D., Pu, H., Sun, D. W., Gao, W. & Xiong, Z. (2014a). Use of Hyperspectral Imaging to Discriminate the Variety and Quality of Rice. *Food Analytical Methods*, **8**, 515-523.

Wang, L., Sun, D. W., Pu, H. & Zhu, Z. (2015). Application of Hyperspectral Imaging to Discriminate the Variety of Maize Seeds. *Food Analytical Methods*, **9**, 1-10.

Wang, W., Heitschmidt, G. W., Ni, X., Windham, W. R., Hawkins, S. & Chu, X. (2014b). Identification of aflatoxin B 1 on maize kernel surfaces using hyperspectral imaging. *Food Control*, **42**, 78-86.

Weinstock, B. A., Janni, J., Hagen, L. & Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (Zea mays L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*, **60**, 9-16.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta*, **653**, 121-130.

Williams, P. C. (1975). Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. *Cereal Chemistry*, **52**, 561-567.

Williams, P. J., Geladi, P., Britz, T. J. & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272-278.

Yang, Y., Chai, R. & He, Y. (2012). Early detection of rice blast (Pyricularia) at seedling stage in nipponbare rice variety using near-infrared hyper-spectral image. *African Journal of Biotechnology*, **11**, 6809-6817.

Yoshizawa, T., Yamashita, A. & Luo, Y. (1994). Fumonisin occurrence in corn from high-and low-risk areas for human esophageal cancer in China. *Applied and Environmental Microbiology*, **60**, 1626-1629.

Zhu, F., Zhang, D., He, Y., Liu, F. & Sun, D. W. (2013). Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen–thawed fish fillets. *Food and Bioprocess Technology*, **6**, 2931-2937.

CHAPTER 3

CHARACTERISATION OF WHITE MAIZE KERNELS USING NIR HYPERSPECTRAL IMAGING

Near infrared hyperspectral imaging with multivariate image analysis was evaluated for its potential to grade whole white maize kernels. The study was based on grading regulations stipulated in South African legislation, with the aim of providing an alternative to the tedious and subjective manual methods currently used. The types of undesirable materials regarded were divided into 13 classes and were imaged with a hyperspectral imaging system with a range 1118 to 2425 nm. Two approaches to data analysis, namely pixel-wise and object-wise analysis, were investigated in the application of PCA and PLS-DA modelling. Two-way classifications were performed to distinguish various types of defective kernels, pinked kernels, yellow maize and foreign matter from sound white maize. The models were developed and validated with independent image data sets containing *ca*. 30 kernels per class. The object-wise PLS-DA models yielded superior results. The pixel-wise classification accuracies ranged 75–99% with no error-free analyses. The object-wise classifications achieved 100% classification accuracy in 8 of the 13 analyses, with the remaining 5 analyses incurring only one error each, thus accuracies ranged 98-100%. Important spectral features were 1219 and 1476 (associated with starch), 1941 (associated with protein), and 2117 (associated with moisture). Results demonstrated good performance in distinguishing between the sound maize class and common grading defects.

INTRODUCTION

Maize is an important cereal crop, providing fundamental food and nutrition as both a human food and animal feed. Maize is a staple food throughout Africa, and in many other parts of the world, such as Latin America and Asia (Johnson, 2000). Grading of maize is an important and critical process, as grade has an influence on the production and quality of the final product, and determines the market value (Serna-Saldivar, 2010). The main aim of grading is thus to facilitate fair commercialisation of maize, and to provide information related to quality for further storage and processing. Currently, South African maize grading practices entail visual assessment of a maize sample to identify and weigh undesirable materials. The grade of the maize consignment is based solely on specified limits of these undesirable materials (e.g. broken kernels, defective kernels and foreign matter) within the sample. The basis of the method is the assumption that the appearance of grain is related to chemical composition, functionality and optimum end use (Serna-Saldivar, 2010). The current grading standards largely relate to storage issues rather than end-use factors (Johnson, 2000). However, with increasing sophistication of the industry, processors have begun to press for a grading system that places greater emphasis on end-use value. This has resulted in many buyers and sellers having agreements that include additional specifications relating to end-use.

The manual grading methods currently in place are tedious and subjective. Considering the importance of accurately determining maize grade, an objective method must be developed that will identify a wide range of defective and foreign materials important to farmers, traders, millers and bakers. Spectral imaging has been identified as a potential alternative to the current methods (Gowen *et al.*, 2007; Amigo *et al.*, 2013). Hyperspectral imaging has been used in many food and agricultural applications, with numerous studies of maize, including hardness prediction (McGoverin and Manley, 2012; Manley *et al.*, 2011; Williams *et al.*, 2009), chemical content prediction (Cogdill *et al.*, 2004; Weinstock *et al.*, 2006), variety identification (Wang *et al.*, 2015) and fungal detection (Del Fiore *et al.*, 2010; Williams *et al.*, 2012). It has also been applied in the quality and safety evaluation of wheat (Mahesh *et al.*, 2015; Manley *et al.*, 2011; Singh *et al.*, 2010), rice (Del Fiore *et al.*, 2010; Wang *et al.*, 2014), and sorghum and barley (McGoverin *et al.*, 2011).

Hyperspectral imaging is the integration of computer imaging and spectroscopy (ElMasry and Sun, 2010). It combines the main features of these two techniques to acquire spatial and spectral information simultaneously. While computer imaging is able to detect surface defects on food products, it is unable to assess intrinsic physical and chemical properties of a sample (Park and Lu, 2015). On the other hand, near infrared (NIR) spectroscopy offers superior detection results compared with conventional imaging, but is unable to pinpoint the location of a defect, a requirement for quantifying the undesirable material content in a graded sample. The respective merits and shortcomings of these two techniques highlight the advantages of combining the two platforms into one analytical system. Spectral imaging is thus highly suited for the study of cereal

products, which show high variability in physical properties and chemical composition within and between kernels.

Hyperspectral imaging generates large datasets, and complex data analysis is required to gain meaningful insight (Esbensen and Geladi, 1989). Multivariate image analysis is often used when processing hyperspectral images. While traditional methods analyse all pixels in an image individually, some applications are better suited for an object-wise approach (Kucheryavskiy, 2013). In this newly developed approach, the average spectrum of all pixels in an object, such as a kernel, is used during modelling. This is an appropriate approach when emulating manual grading, as whole maize kernels must be identified as either sound or defective. There is no value in quantifying the number of defective pixels, for instance during pixel-wise analysis, because any presence of the defect on a kernel is unacceptable. Each kernel should rather be observed as the lowest unit of measurement, as in object-wise analysis.

The aim of the study was to evaluate the capability of NIR hyperspectral imaging to separate sound white maize kernels from common undesirable material types encountered in the South African maize industry. Two approaches to data analysis, namely pixel-wise and object-wise data analysis, will be utilised and compared.

MATERIALS AND METHODS

Samples

Maize kernels and undesirable materials were obtained from the South African Grain Laboratory (SAGL, Pretoria, South Africa) and were graded visually by expert graders at the SAGL according to South African grading regulations (Department of Agriculture, 2009). The 13 most prolific undesirable materials encountered in the South African maize industry were selected as classes to be distinguished from the sound maize class (Fig. 4.1). Classes included defective white maize (heat damage, water damage, screenings/broken kernels, *Fusarium* fungal damage and *Diplodia* fungal damage); pinked white maize; other colour maize (i.e. yellow maize); and foreign materials (wheat, soy, sunflower seeds, sorghum and maize plant material). The classes relate to the South African grading legislation, which outlines 4 main categories. Either the main category or several sub-categories were used as classes.

Figure 3.1 Digital image of (a) sound white maize and the 13 undesirable materials, i.e. (b) *Fusarium* damage; (c) *Diplodia* damage; (d) pinked maize; (e) water damage; (f) rodent damage; (g) heat damage; (h) plant material; (i) screenings; (j) wheat; (k) sorghum; (l) soy; (m) sunflower; and (n) yellow maize.

NIR hyperspectral system

Hyperspectral images were acquired using a Xenics short wave infrared (SWIR) XEVA camera with an ImSpector N25E spectrograph that included a Mercury–Cadmium–Telluride (HgCdTe) detector (SPECIM Ltd, Oulu, Finland). The samples were carried on a conveyer belt (Burgermetrics SIA, Riga, Latvia) during acquisition on the pushbroom instrument. The system was controlled with HyperPro software (BurgerMetrics SIA, Riga, Latvia). Individual images were acquired within a spectral range of 1118 to 2425 nm at 6.3 nm spectral resolution between the 209 spectral points. The frame rate was 100 Hz and the exposure time ranged 0.8–1.2 ms. Images were 320 pixels wide, with a varying length of *ca.* 400 pixels. White and dark references were captured prior to each sample image and were subsequently used for image correction and calibration. A 100% reflectance standard was used for the white reference, and the shutter was closed for the dark reference.

Image acquisition

For the calibration set, images of the sound maize class and each of the 13 undesirable material classes were captured individually. These data sets consisted of 30 kernels/objects of each class, with two exceptions, where 24 pinked kernels were used due to limited availability and 60 screenings pieces due to their small size. For the classes of maize kernels (sound maize; *Fusarium* damage; *Diplodia* damage; water damage; heat damage; rodent damage; and yellow maize), the

kernels were first imaged with the germ facing upwards (towards the camera), and a second time with the germ facing downwards (away from the camera). Thus the germ-up and -down oriented versions of the images consisted of the same kernels in the same positions. The foreign matter classes (plant material, wheat, sorghum, soy and sunflower seeds) were only imaged once, as the foreign matter often had no obvious germ. The validation set was acquired in the same manner on new sets of kernels and objects (same number as in calibration set) for each of the 13 undesirable material classes.

Hyperspectral image analysis

Images were analysed using the Evince v.2.7.0 (UmBio AB, Umeå, Sweden) spectral image analysis software package. The germ-up and -down images for each pair (sound and undesirable materials) were mosaiced and each image mosaic was analysed individually. A summarised illustration of the image analysis process is given in Fig. 3.2.

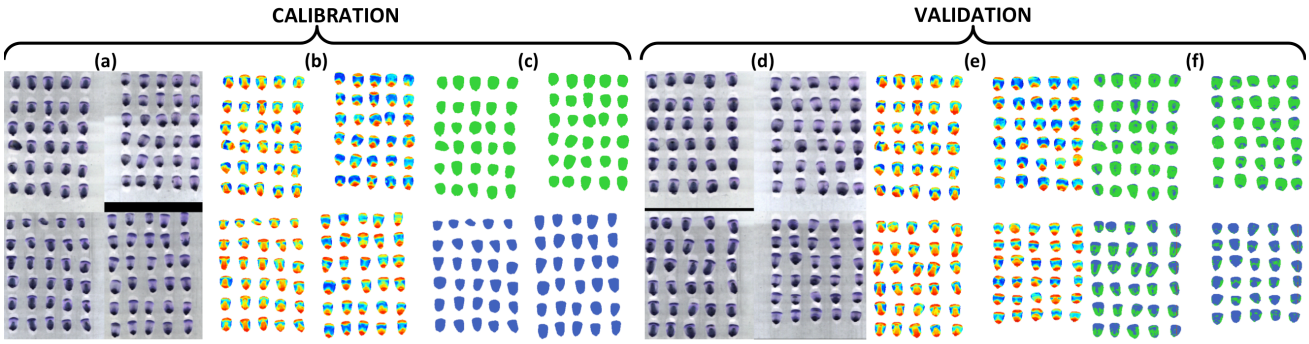

Figure 3.2 Summary of image analysis process, illustrated using sound class (top) vs. yellow class (bottom) with the pixel-wise approach (object-wise analysis followed the same method). The left half of each section (a–f) is germ-up, and the right half is germ-down. Each half contains the same kernels in the same positions, to allow for inspection of both sides. (a) Raw mosaic image (at 1426 nm) of calibration samples arranged in known order of classes; (b) PCA score image of calibration samples after pre-processing; (c) Samples were assigned classes (sound as green and defect as blue) and PLS-DA model was calibrated; (d) Raw mosaic image (at 1426 nm) of independent validation samples arranged in known order of classes; (e) PCA score image of validation samples after pre-processing; and (f) PLS-DA model applied to generate classification image.

Image correction and cleaning

The image calibration and correction from reflectance to pseudo-absorbance was done automatically in the Evince software package according to Eq. 3.1.

PCA was applied to the mean-centred absorbance mosaic images. Score plots and score images were used interactively to identify unwanted pixels, e.g. outliers, sample stage background, dead pixels, shading errors and edge effects, which were subsequently removed from the dataset (Esbensen and Geladi, 1989). The cleaned image was used in subsequent analysis. PCA was recalculated with additional components after obtaining a cleaned image.

$$I_{\lambda,n} = -log_{10}\left[\left(\frac{S_{\lambda,n} - B_{\lambda,n}}{W_{\lambda,n} - B_{\lambda,n}}\right)\right] \quad (3.1)$$

Where:

n = Pixel index variable $(n = 1 \ldots N)$ of the reorganised hypercube

$I_{\lambda,n}$ = Standardised absorbance intensity, pixel n, at wavelength λ

$S_{\lambda,n}$ = Sample image, pixel n, at wavelength λ

$B_{\lambda,n}$ = Dark reference image, pixel n, at wavelength λ

$W_{\lambda,n}$ = White reference image, pixel n, at wavelength λ

Principal component analysis

After various pre-processing treatments were evaluated, Savitsky-Golay transformation (2nd order polynomial, 1st derivative, 15 points) and standard normal variate (SNV) transformation was applied to the data (Rinnan *et al.*, 2009). In addition, noisy wavelengths 2218 to 2425 nm were removed. Data were analysed with pixel-wise and object-wise PCA (Kucheryavskiy, 2013), and the PC score images, score plots and loading line plots were examined.

Partial least squares discriminant analysis

PLS-DA models were calculated to distinguish between the two classes, namely sound and respective undesirable material. The aim of PLS-DA is to model the relationship between a set of predictor variables (**X**), namely the spectral data, and a set of responses (**Y**) (Næs *et al.*, 2002). The response was the assignment of an object to a mutually exclusive class, sound or undesirable material. Pixel-wise and object-wise PLS-DA models were calculated on the calibration image data, with no cross-validation for the pixel-wise models, as the number of variables was too large, and full cross-validation for the object-wise models. The mean spectrum of all pixels in an object was used during object-wise analyses. The PLS-DA models were applied to the validation image data, and a classification image was generated, referred to as the unaltered classification image in later sections.

The correct and incorrect predictions with regards to the kernel's assigned classes were totalled for the pixel-wise and object-wise classifications using two different methods. Due to the large number and small size of the pixels, the pixel-wise predictions were totalled across both the germ-up and -down images as an average percentage of the total pixels. The larger size and manageable number of the objects in the object-wise predictions allowed for comparison of the results for individual kernels' germ-up and -down classifications. Correct and incorrect object-wise predictions, with reference to the kernel's assigned classes, were totalled for each PLS-DA classification image, as illustrated in Fig. 3.3. For both approaches, if the overall classification was correct, it was counted as a correct classification, forming part of the correct classification accuracy (%). If it was incorrect,

it was classified as either a false positive or false negative, forming part of the false positive error (%) or false negative error (%), respectively.

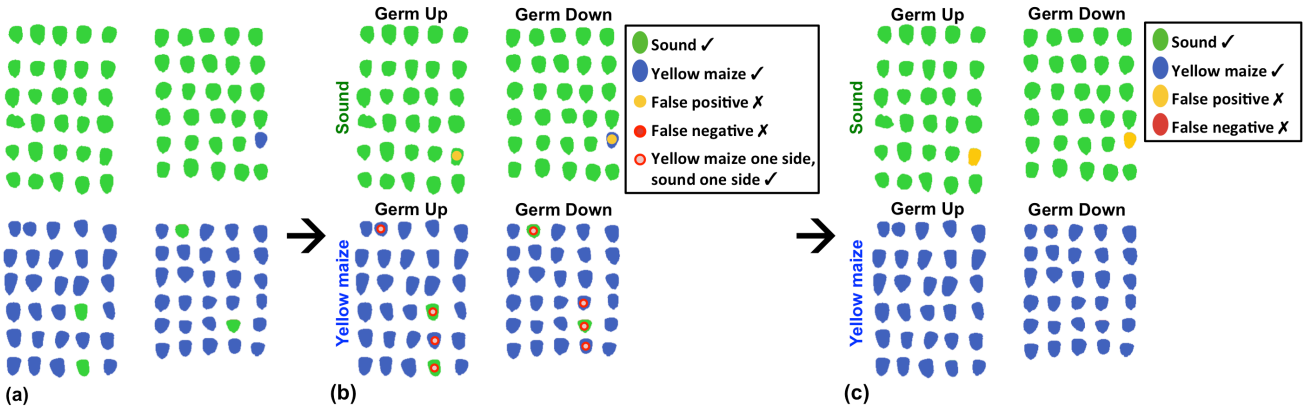

Figure 3.3 Method used for totalling the object-wise classification results, illustrated using yellow maize class vs. sound class. (a) Unaltered object-wise classification image; (b) Assignment of overall classification; and (c) Overall object-wise classification image.

Fig. 3.3a displays the unaltered classification image after external validation. It is called the unaltered image because this is only the model output before human evaluation of the results. It was difficult to interpret because the results indicate assigned class with no reference to known class. Fig. 3.3b illustrates the process of determining the overall classifications of kernels. Both the germ-up and -down unaltered classification images were inspected to evaluate if the two sides of a kernel have the same or different classification. The kernel's overall classification was assigned according to Table 3.1. The overall classification of the kernel was deemed correct or incorrect (false) with reference to the kernel's known class. Fig. 3.3c is an overall classification image, which displays only the overall classification for clarity. By this stage, the germ-up and -down sections of the image were identical, as each kernel received one overall classification.

Germ-up Classification	Germ-down Classification	Overall Classification
Sound	Sound	Sound
Sound	Defect	Defect
Defect	Sound	Defect
Defect	Defect	Defect

Table 3.1 Method used to determine the overall classification of a kernel based on the classification image generated from the germ-up and -down image data

Classification accuracy, false positive error, and false negative error were calculated according to Eq. 3.2, 3.3 and 3.4, respectively. An undesirable material classification was a positive response and a sound classification was a negative response. A false positive occurred when a sound kernel was incorrectly classified as undesirable material, and a false negative was when an

undesirable material was incorrectly classified sound.

$$\text{Classification accuracy (\%)} = \frac{\text{Correct Defect + Correct Sound}}{\text{Total}} \times 100\% \quad (3.2)$$

$$\text{False positive error (\%)} = \frac{\text{False Positives}}{\text{Total}} \times 100\% \quad (3.3)$$

$$\text{False negative error (\%)} = \frac{\text{False Negatives}}{\text{Total}} \times 100\% \quad (3.4)$$

Where:

Correct Defect = Correctly classified class 'Defect' kernels

Correct Sound = Correctly classified class 'Sound' kernels

False Positives = 'Sound' class kernels incorrectly classified 'Defect' class

False Negatives = 'Defect' class kernels incorrectly classified 'Sound' class

Total = Sum of 'Defect' and 'Sound' kernels

Sensitivity and Specificity were calculated according to Eq. 3.5 and 3.6, respectively. The sensitivity describes the probability that undesirable materials will be detected and correctly classified, and is sometimes described as the true positive rate. Specificity is the probability that a sound kernel will be classified correctly, also known as the true negative rate.

$$\text{Sensitivity (\%)} = \frac{\text{Correct Defect}}{\text{Total Defect}} \times 100\% \quad (3.5)$$

$$\text{Specificity (\%)} = \frac{\text{Correct Sound}}{\text{Total Sound}} \times 100\% \quad (3.6)$$

Where:

Correct Defect = Correctly classified class 'Defect' kernels

Correct Sound = Correctly classified class 'Sound' kernels

Total defect = Total 'Defect' kernels (Correct Defect + False Negatives)

Total Sound = Total 'Sound' kernels (Correct Sound + False Positives)

RESULTS AND DISCUSSION

Spectral Analysis

Spectral analysis in the NIR region is a widely used technique for evaluating cereal properties. The spectra generated by the hyperspectral image acquisition were inspected to investigate the chemical properties of the samples. The raw average pseudo-absorbance spectra (1118 to 2211 nm) of the sound white maize, defective white maize and pinked white maize classes are given in Fig. 3.4, and sound white maize, yellow maize and the foreign matter classes are given in Fig. 3.5. The spectra of the classes tend to follow a similar form, but varied in relative intensity. The only exception was the first section of the sunflower seeds' spectrum. This is likely due to the seed being contained within an outer husk. The NIR wavelengths are not able to fully penetrate this husk, and thus the inner seed was not able to interact fully. The outer husk has a very different chemical

signature to the other cereal commodities. The differences in intensity exhibited between the spectra of the classes cannot be related to the internal chemical composition of the kernels directly, as the given spectra are unprocessed. Instead, the varying intensities are likely associated to physical effects such as scattering. Furthermore, direct interpretation of heterogeneous samples like whole maize kernels are difficult, as the spectra contain the combined information from all kernel components. However, general trends were observed. Four main absorption bands in the spectra of the 14 classes were exhibited at 1219, 1476, 1941 and 2117 nm. The peaks at 1219 (C–H stretch second overtone) and 1476 nm (O–H stretch first overtone) indicate variation in the starch composition of the endosperm (Delwiche and Hareland, 2004; Manley *et al.*, 2009). The broad peak at 1476 nm also has a contribution from moisture around 1430 nm. A clear moisture peak appeared due to O–H stretch at 1941 nm (Manley *et al.*, 2009) and the peak present at 2117 nm relates to amino acids (N–H stretch first overtone), and thus protein content (Fernández-Ibañez *et al.*, 2009).

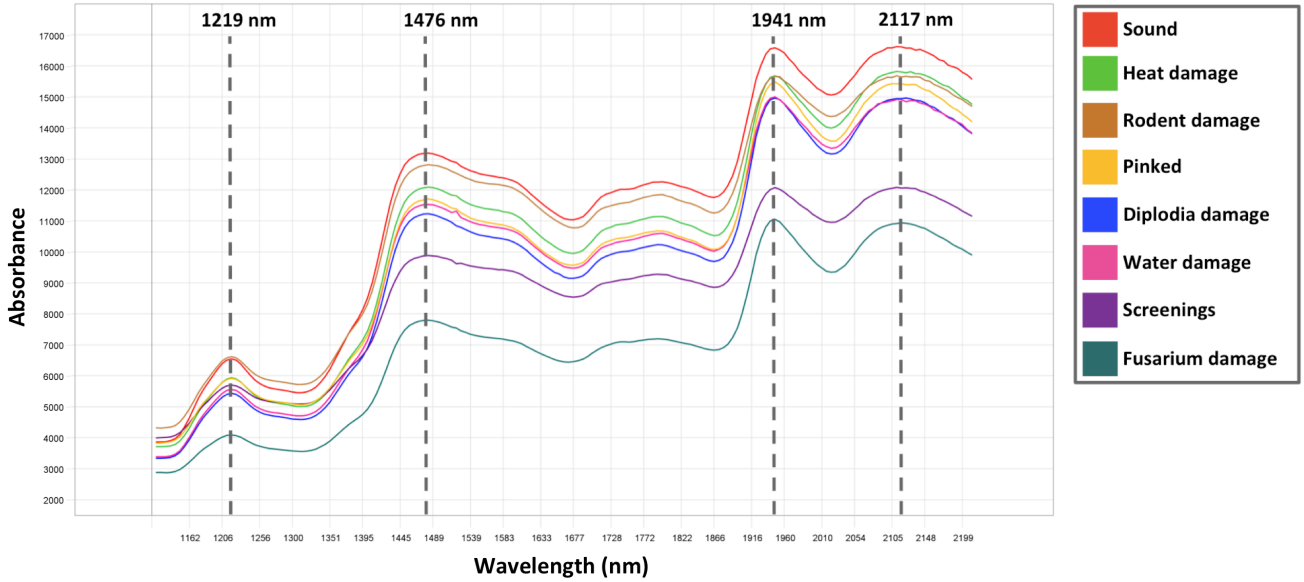

Figure 3.4 Unprocessed average pseudo-absorbance spectra for sound white maize; white maize defects (heat damage, rodent damage, *Diplodia* damage, water damage, screenings and *Fusarium* damage); and pinked white maize.

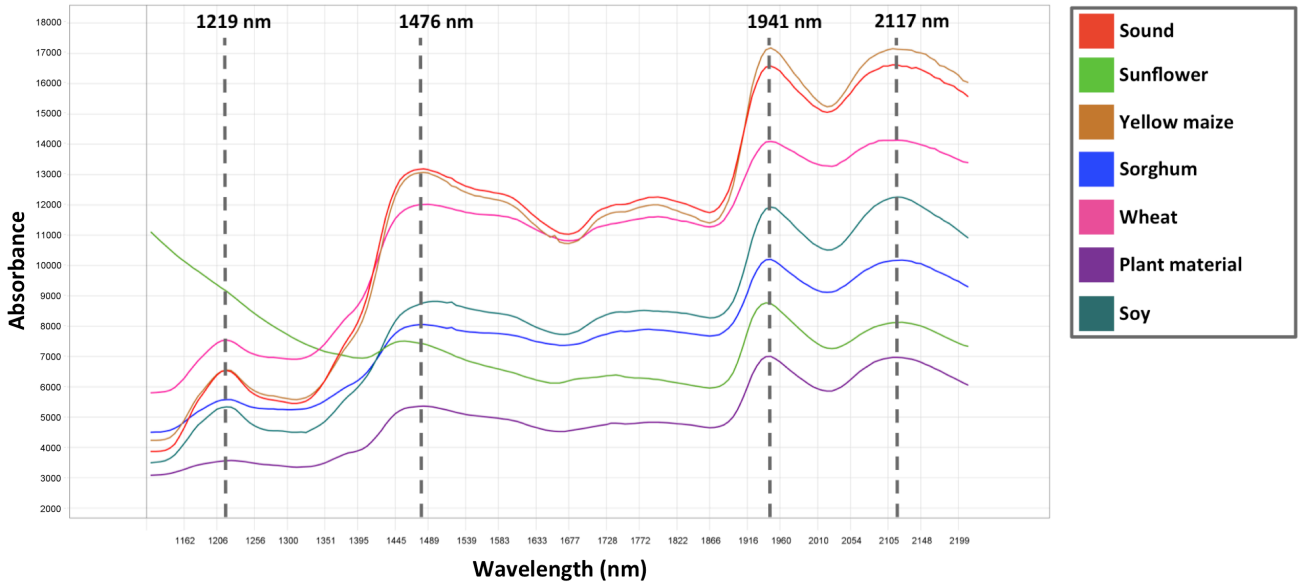

Figure 3.5 Unprocessed average pseudo-absorbance spectra for sound white maize; yellow maize; and foreign matter (sunflower seeds, sorghum, wheat, plant material and soy).

Multivariate Data Analysis

Principal Component Analysis

PCA was calculated to examine the qualitative difference between sound white maize and the various undesirable material classes in the principal component (PC) space. The classes heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are used to illustrate the PCA results, as given in Fig. 3.6 to 3.11 for the pixel-wise analyses and Fig. 3.12 to 3.17 for the object-wise analyses. Note that the scores plots (a) are shaded according to density of points, while the scores images (b) are shaded according to scores value where blue is negative and red is positive.

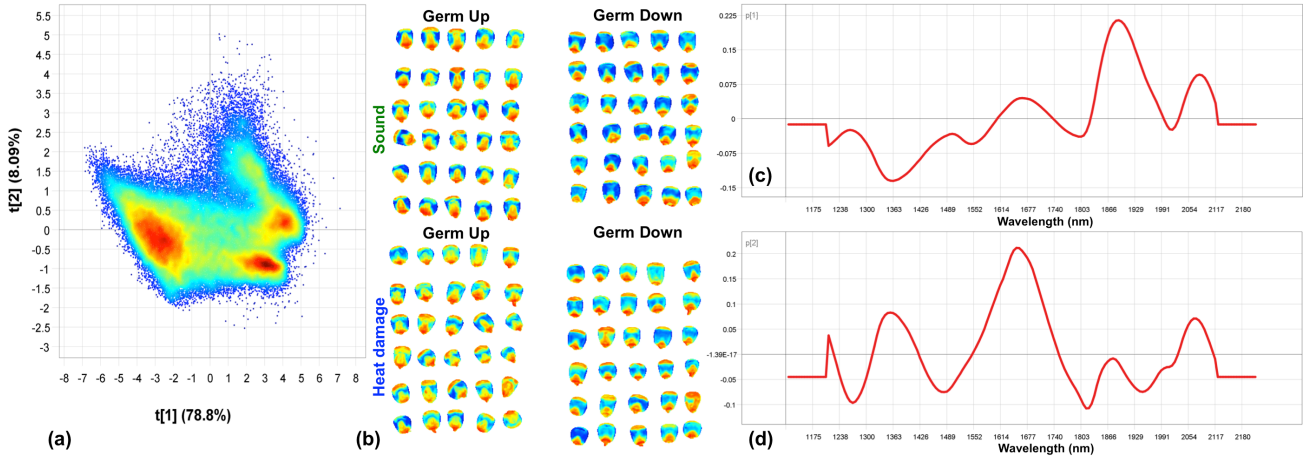

Figure 3.6 Pixel-wise PCA analysis of heat damage class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (79% SS) vs. PC2 (8% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

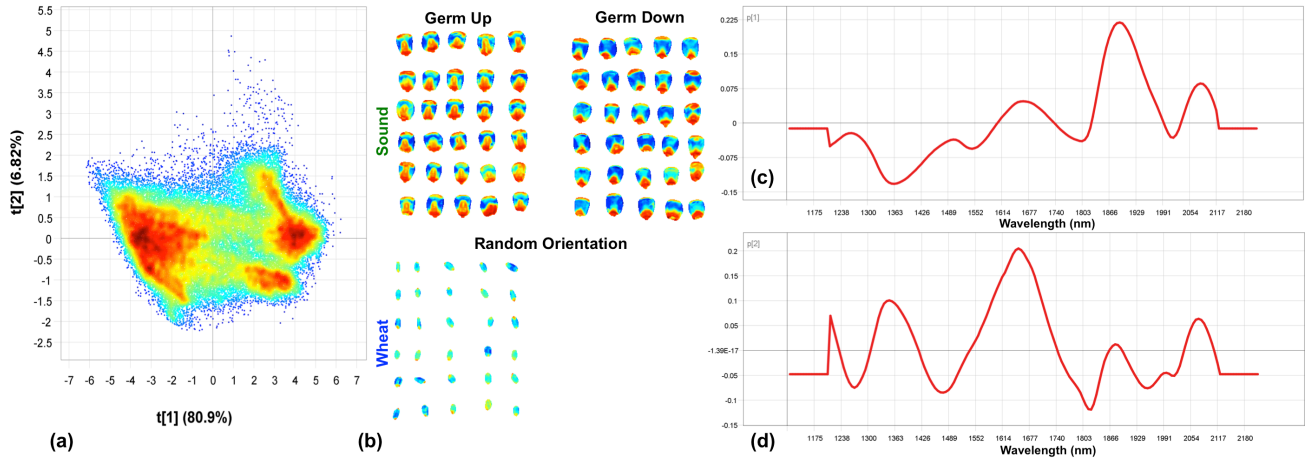

Figure 3.7 Pixel-wise PCA analysis of wheat class vs. sound class. Slight separation of classes was observed. Scores given as (a) PCA score plot of PC1 (81% SS) vs. PC2 (7% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

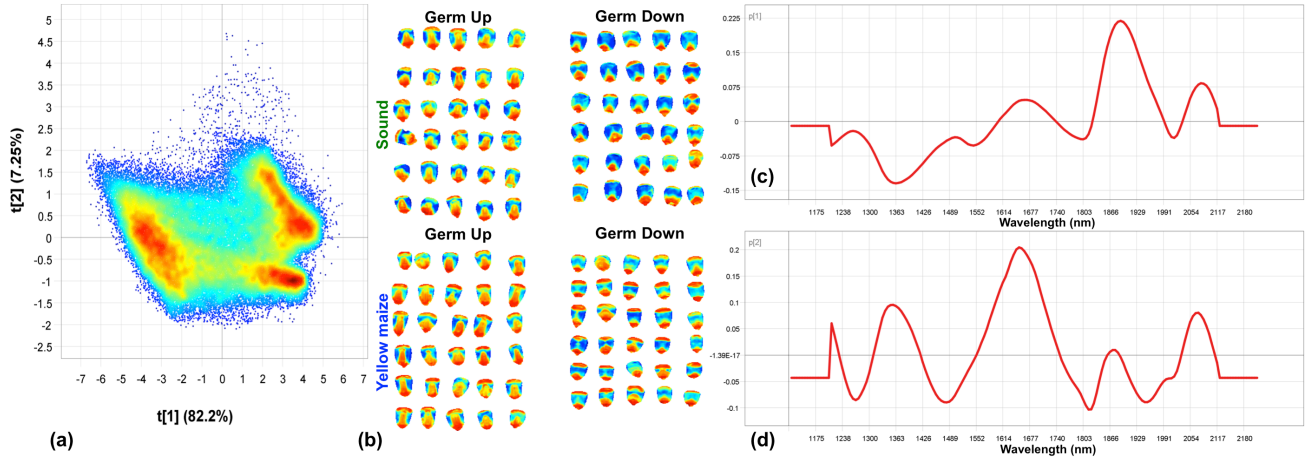

Figure 3.8 Pixel-wise PCA analysis of yellow maize class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (82% SS) vs. PC2 (7% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

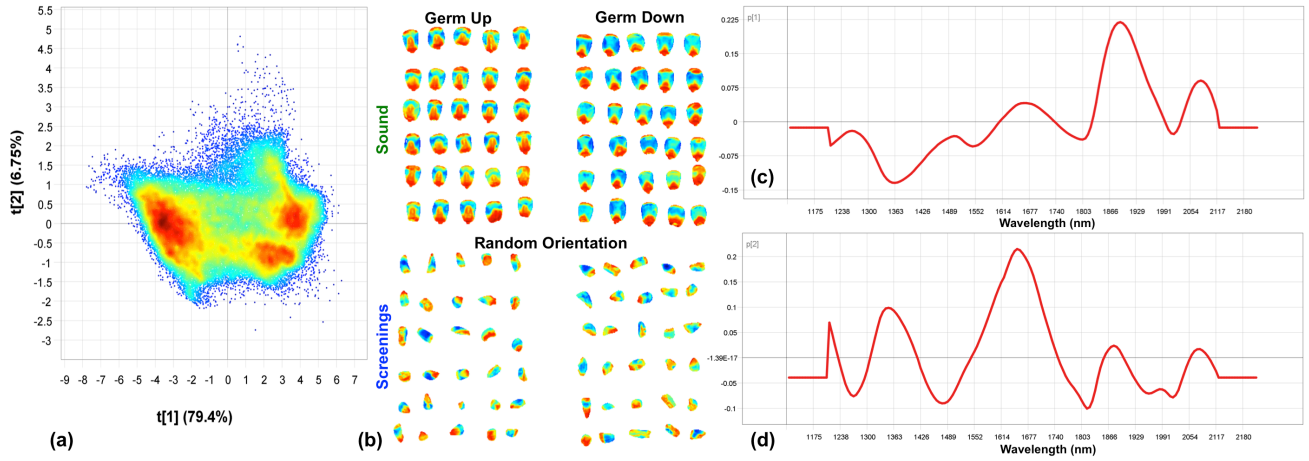

Figure 3.9 Pixel-wise PCA analysis of screenings class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (79% SS) vs. PC2 (6% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

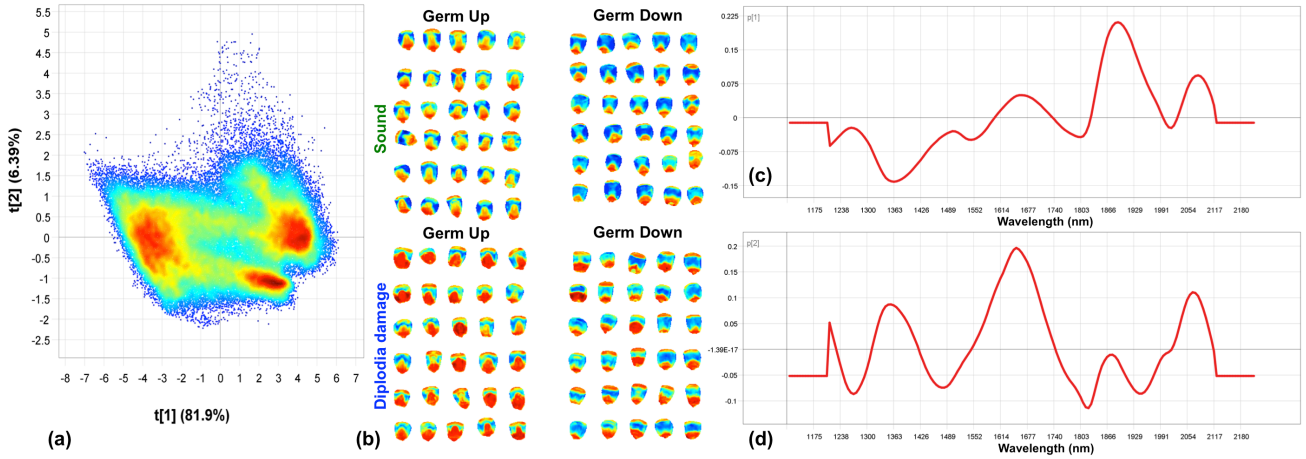

Figure 3.10 Pixel-wise PCA analysis of *Diplodia* damage class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (82% SS) vs. PC2 (6% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

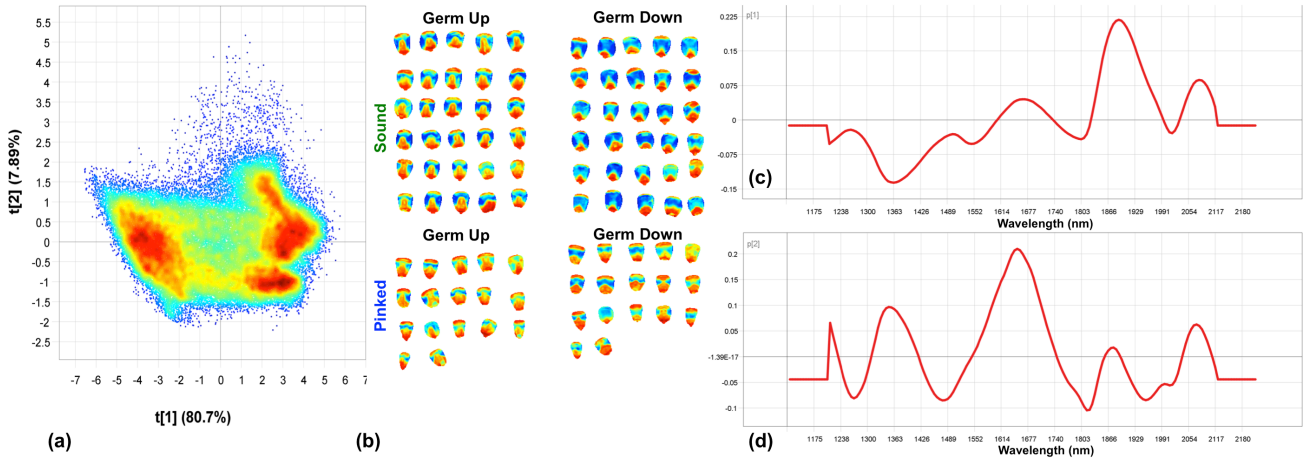

Figure 3.11 Pixel-wise PCA analysis of pinked class vs. sound class. Minimal separation of classes was observed. Scores given as (a) PCA score plot of PC1 (81% SS) vs. PC2 (8% SS); and (b) PCA score image (PC1). Loadings line plots given for (c) PC1; and (d) PC2.

During the pixel-wise PCA analyses the spectrum from each pixel in the cleaned image is included during model calculation. The PCA score plots for PC1 vs. PC2 showed a major overlap of the two classes (Fig. 3.6a–3.11a). Due to the density and number of points on the plot, there were no visible areas associated with a single class. The variance accounted by PC1 was likely due to variation within the kernels and their anatomical kernel components, as shown in the score plots and score images (Fig. 3.6b–3.11b). The clusters that appeared in the direction of PC1 in the score plots were found to correspond with specific components of the kernels in the PC1 score images, such as soft endosperm, hard endosperm and germ, regardless of the kernel's class. For the maize classes heat damage (Fig. 3.6), water damage, rodent damage, yellow maize (Fig. 3.8), screenings (Fig. 3.9), *Fusarium* damage, *Diplodia* damage (Fig. 3.10) and pinked kernels (Fig. 3.11), variance in PC2 continued to arise from within the kernel. PC3 (not shown) accounted for the difference in germ orientation, where a cluster appeared in the score plot related to the large germ appearing on the

germ-up oriented kernels. The differences between the classes often only appeared to became a source of variation at PC4 and PC5 (not shown), although the observable differences in the plots remained slight. The foreign matter classes exhibited differences between the two classes in earlier PCs. Wheat (Fig. 3.7), sorghum and soy exhibited similar variance within kernels in PC1, but exhibited clear separation of classes from PC2 onward. This was likely due to less similarity between chemical components of these cereal commodities. Sunflower seeds (which included the outer husk) and plant material had extremely different chemical surface characteristics to maize, and thus were clearly separated in PC1.

The lack of separation between the classes in the pixel-wise PCA results was observed in the colour gradients of the score images (Fig. 3.6b–3.11b). The colour shading indicates the score value of each pixel. A positive PC1 score value was displayed as hot colour shading while a negative score value was displayed as cool colour shading. One may easily observe that in many of the analyses, such as heat damage (Fig. 3.6b), each kernel displays the full complment of colours, confirming that the variance related to PC1 was within kernels due to variation of their constituents. However, in the cases of sunflower seeds and plant material, a severe divide appeared. The colour shading of the maize in the score image had little to no variation, as the differences within each class paled in comparison to the large difference between the classes. This type of separation should have ideally occurred in more of the 13 analyses, but was prevented due to similarities between classes. Overall, the overwhelming lack of separation between sound maize and most undesirable materials emphasises the need to investigate an object-wise approach, where these internal variances were excluded by averaging the spectra of all individual pixel spectra in a single kernel.

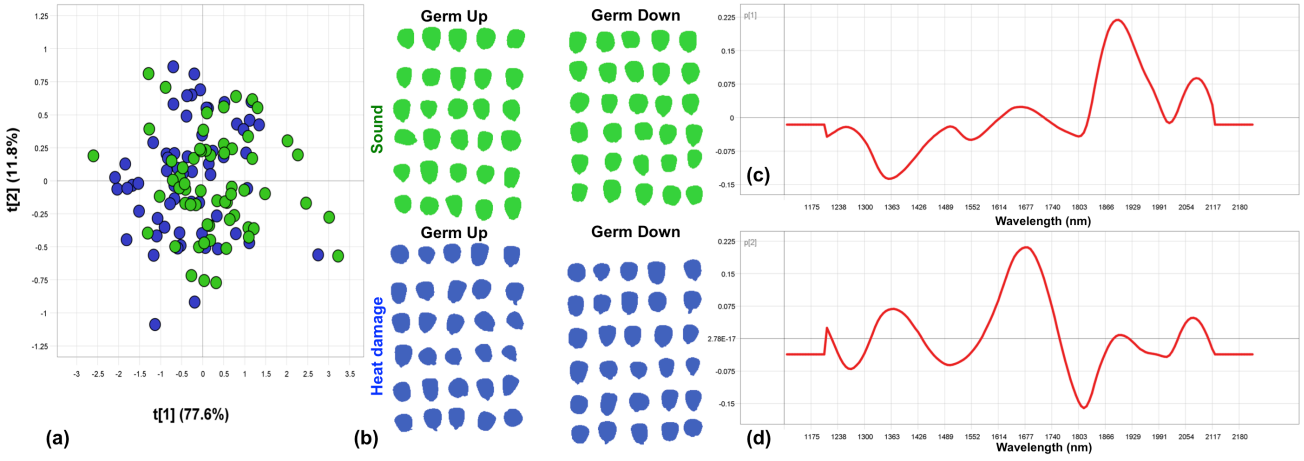

Figure 3.12 Object-wise PCA analysis of heat damage class vs. sound class. Minimal separation of classes was observed. (a) PCA score plot of PC1 (78% SS) vs. PC2 (12% SS). (b) Classes of sound (green) and heat damaged (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

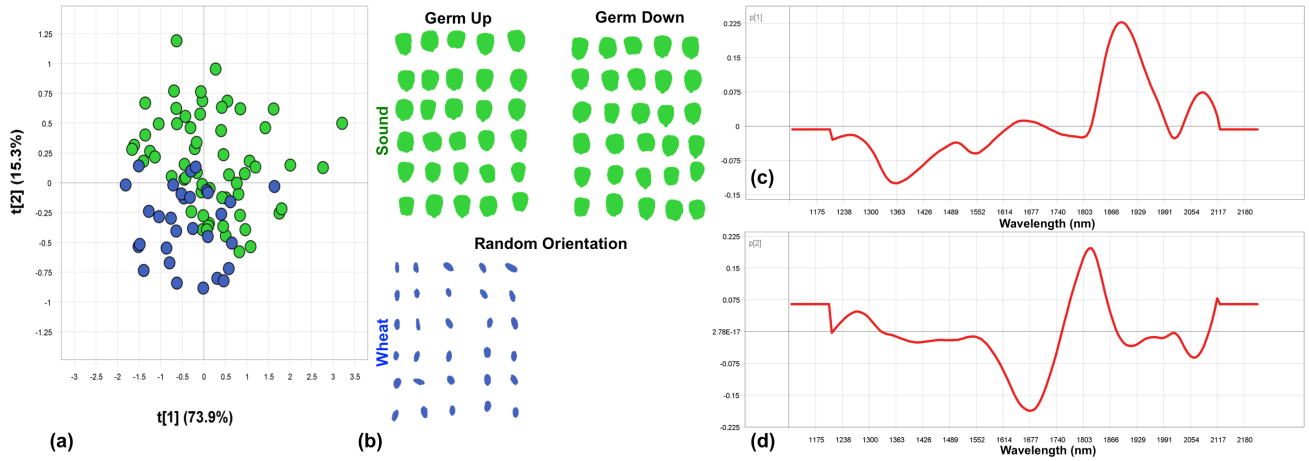

Figure 3.13 Object-wise PCA analysis of wheat class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (74% SS) vs. PC2(15% SS). (b) Classes of sound (green) and wheat (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

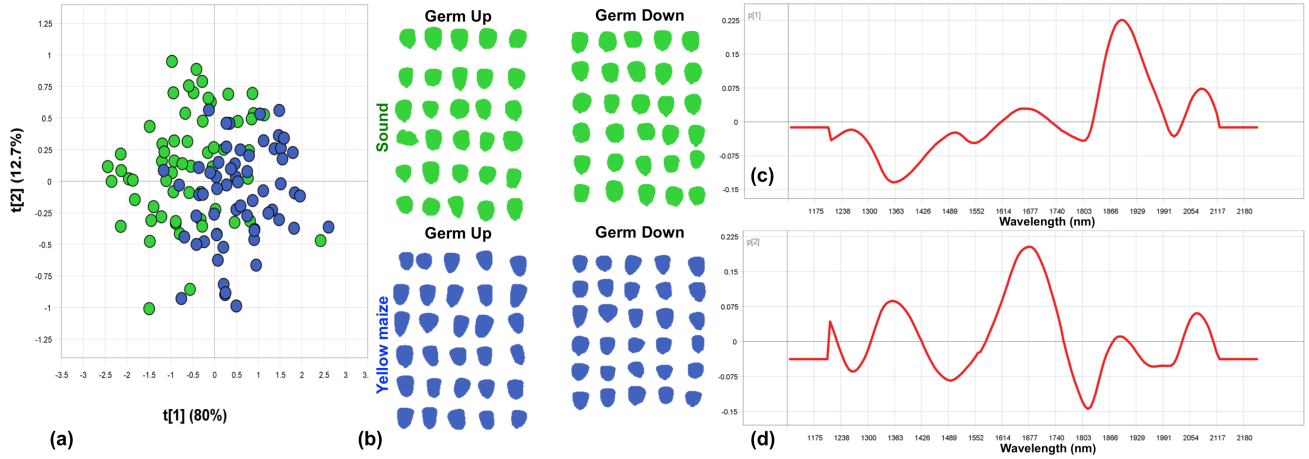

Figure 3.14 Object-wise PCA analysis of yellow maize class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (80% SS) vs. PC2 (13% SS). (b) Classes of sound (green) and yellow maize (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

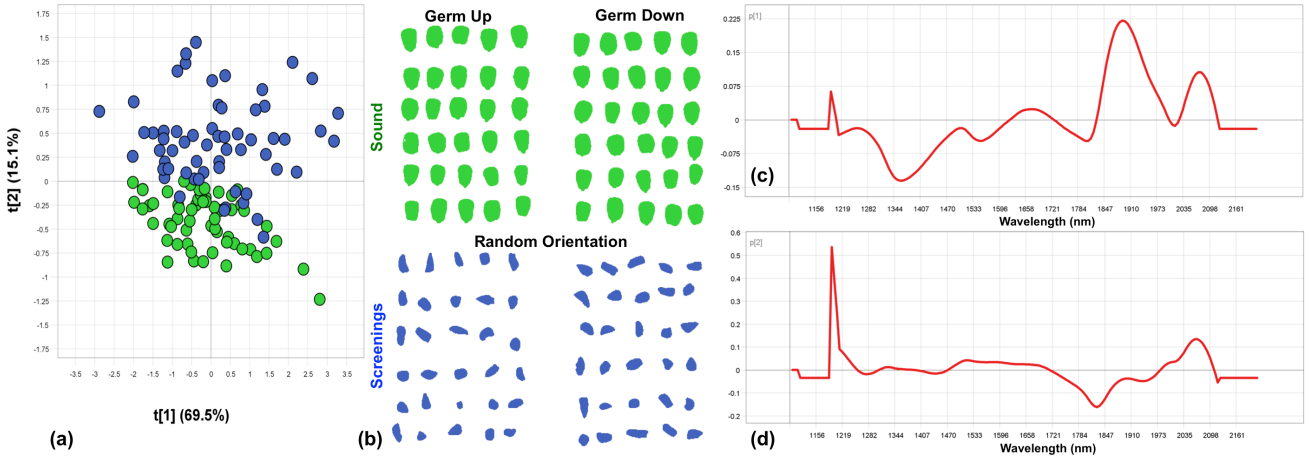

Figure 3.15 Object-wise PCA analysis of screenings class vs. sound class. Good separation of classes was observed. (a) PCA score plot of PC1 (70% SS) vs. PC2 (15% SS). (b) Classes of sound (green) and screenings (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

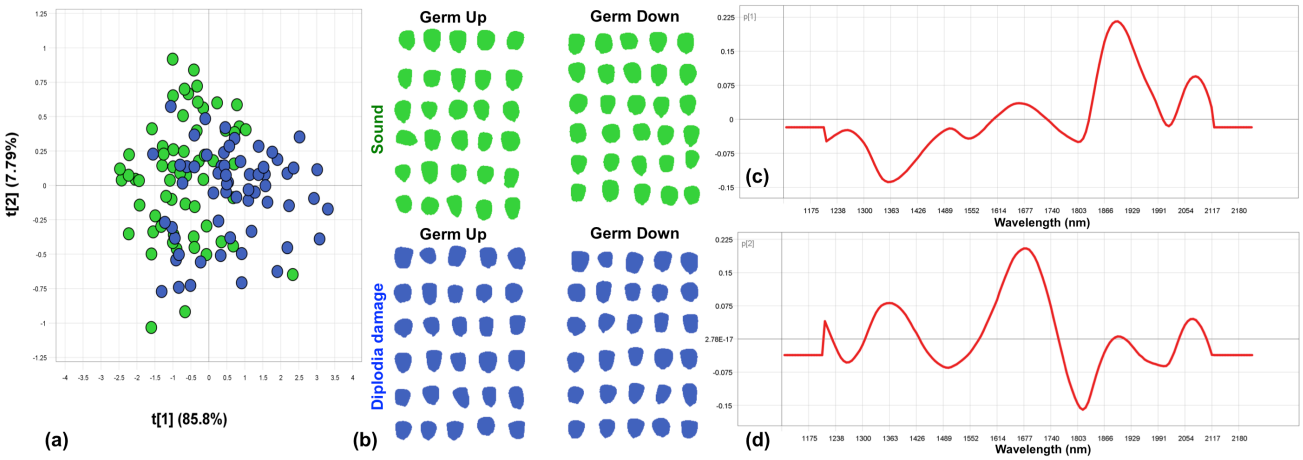

Figure 3.16 Object-wise PCA analysis of *Diplodia* damage class vs. sound class. Slight separation of classes observed. (a) PCA score plot of PC1 (86% SS) vs. PC2 (8% SS). (b) Classes of sound (green) and *Diplodia* damage (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

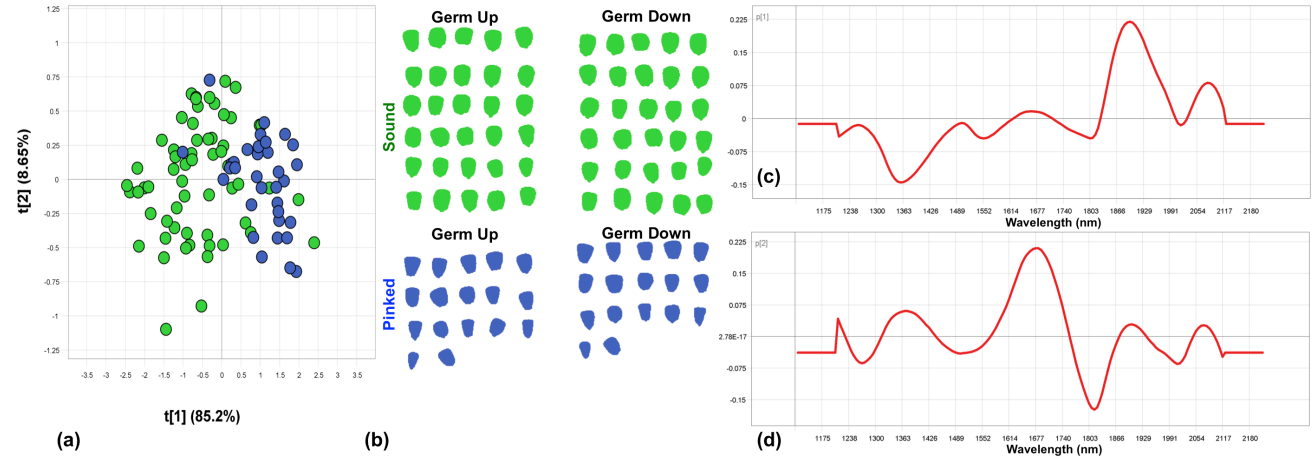

Figure 3.17 Object-wise PCA analysis of pinked class vs. sound class. Fair separation of classes was observed. (a) PCA score plot of PC1 (85% SS) vs. PC2 (9% SS). (b) Classes of sound (green) and pinked (blue) objects. Loadings line plots given for (c) PC1; and (d) PC2.

The object-wise PCA analyses revealed distinctions between the classes in PC1 that were not easily discerned in the pixel-wise results. The PCA score plots of PC1 vs. PC2 (Fig. 3.12a–3.17a) of most classes show a clear separation of points between the sound (green) and undesirable material (blue) objects. The score plot of the screenings class (Fig. 3.15a) exhibited the most distinct separation that occurred predominantly in the direction of PC2 with two clearly defined clusters. Similarly, wheat (Fig. 3.13a) exhibited more distinct separation due to variance in PC2 as opposed to PC1. The variance explained in PC1 may relate to intra-class separation (e.g. hardness effects), while PC2 accounted for the differences between the two classes. The remaining classes, as illustrated by heat damage (Fig. 3.12), yellow maize (Fig. 3.14), *Diplodia* damage (Fig. 3.16) and pinked kernels (Fig. 3.17) were separated predominantly in the direction of PC1, all with the sound class being associated with negative scores and the undesirable material class with positive scores.

The heat damage class showed the lowest degree of separation and an overlap of most points, likely due to the similarity of the sound maize and the less severe heat damage defect. One drawback of performing an object-wise PCA analysis is that there is no intra-kernel spatial spectral information to explore. The number of points in the pixel-wise analyses gives improved insight into the chemical distribution of the samples. Although this learned information may not be used directly for the calculation of further classification models, the preliminary exploratory analysis allowed by pixel-wise PCA aids in brushing and optimising pre-processing techniques. Thus, pixel-wise PCA is recommended when exploring the image dataset, while it appears that clearer separation of samples may be achieved through an object-wise approach.

Although the spectral signatures of the sound and undesirable material classes were very similar, there are numerous physical and chemical components that could vary and give rise to slight spectral differences. These differences were most probably due to the increase and decrease in concentrations of various metabolites as a result of defect onset. To study the relationship between the samples' spectra and the PCA model, one must refer to the loadings. The loadings can be interpreted as the relative weightings of the spectral variables in each PC. The loadings line plots of a PC shows how each variable contributed to the model (Esbensen *et al.*, 2002). These provide valuable information, but are especially informative when examined in conjunction with the score plots (Fig. 3.6-3.17), as information about the variables and objects are best studied together. A variable with positive loadings contributes to positive PC scores, and negative loadings contribute to negative PC scores. In other words, objects on the positive side of the score plot will have exhibited high spectral responses (pseudo-absorbance) for the variables on the positive side of the loading line plot, and vise versa (Esbensen *et al.*, 2002).

The loading line plots of PC1 and PC2 (Fig. 3.6c–3.17c and 3.6d–3.17d) were similar throughout the classes in both the pixel-wise and object-wise results. Prominent positive and negative peaks were observed in PC1 at 1260, 1500, 1640, 1890 and 2080 nm, and 1363, 1540, 1803 and 2010 nm, respectively. The peaks in the loading line of PC2 were located at the same spectral variables, although occasionally appeared in an inverted orientation. The loadings lines represent data that had been treated with the Savitsky-Golay first derivative transformation, with a second order polynomial and a filter width of 31 points as parameters. This pre-processing transformation is known to cause a shift in the baseline of the curve (Mark and Workman Jr, 2003; Tsai and Philpot, 1998). By calculating the first derivative with relatively large spacing, the wavelength corresponding to the maximum of the observed peak may shift. In this study, the phenomenon was caused by the parameter of 15 points spacing on either side. As a result, the loadings line has been condensed into the space after the first 15 points (1206 nm) and before the last 15 points (2117 nm). One can observe a flat line before and after these points in the given plots. The range was effectively reduced from 1118–2211 nm to 1206–2117 nm, however all peaks occurring in the original range were retained and shifted closer together. The shorter wavelengths

(1206–1658 nm) were shifted to the right and the longer wavelengths (1665–2117 nm) were shifted to the left. Logically, a peak near to the end-points of the loadings line will undergo a more severe shift effect, while the centre point remains unshifted. The strong moisture peak observed during the spectral analysis of the raw spectral data corresponded to 1945 nm. A clear demonstration of this shift was evident in the results, where the loadings line plots of all classes exhibited a peak at 1890 nm, likely corresponding with moisture. The peak thus exhibits a shift of 55 nm to the left. Furthermore, by calculating the first derivative, the original sharp single peaks were split into two peaks, one positive and one negative. The first derivative spectrum will cross zero at the original (raw) spectrum's maximum, with the two peaks either side exhibiting maxima where the original peak had its maximum gradients. The moisture peak was thus likely to be represented by a pair of peaks, and the loadings line peak at 1803 nm was thus also related to moisture content of the samples. The loadings line peaks at 1260 nm (positive) and 1363 nm (negative) were shifted right and correspond to the first starch content related peak identified in the spectral analysis at 1220 nm, and the peaks at 1500 nm (positive) and 1540 nm (negative) correspond with the second starch peak identified at 1476 nm. The loadings line peaks at 2010 nm (negative) and 2080 nm (positive) represent a left-shifted protein peak originally identified at 2117 nm.

Partial Least Squares Discriminant Analysis

The results of both image data analysis approaches are given in Tables 3.2 and 3.3. The coefficient of determination (R^2) allows evaluation of the fit of the classification model to the calibration data, while the cross-validated coefficient of determination (Q^2) was calculated for the object-wise data based on full cross-validation. Full cross-validation of the pixel-wise models was not viable due to the large size of the datasets. The parameters classification accuracy, false positive error, false negative error, specificity and sensitivity indicate the quality of the classification results.

None of the classifications by pixel-wise PLS-DA achieved 100% accuracy, and varied widely between 62.94 (pinked white maize) and 99.65% (sunflower seeds) (Table 3.2). All of the object-wise PLS-DA models performed well (Table 3.5) with 8 of the 13 analyses achieving 100% classification accuracy, including: *Diplodia* damage; heat damage; screenings; plant material; wheat; sorghum; soy; and sunflower. The 5 remaining classes only predicted one kernel incorrectly. The classes with one false positive (a sound kernel was incorrectly classified as undesirable material), include *Fusarium* damage and yellow maize, and the classes with one false negative (undesirable material was not detected) included water damage, rodent damage and pinked white maize kernels.

The classes heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are used to illustrate the PLS-DA results, as given in Fig. 3.18 to 3.23 for the pixel-wise analyses and Fig. 3.24 to 3.29 for the object-wise analyses.

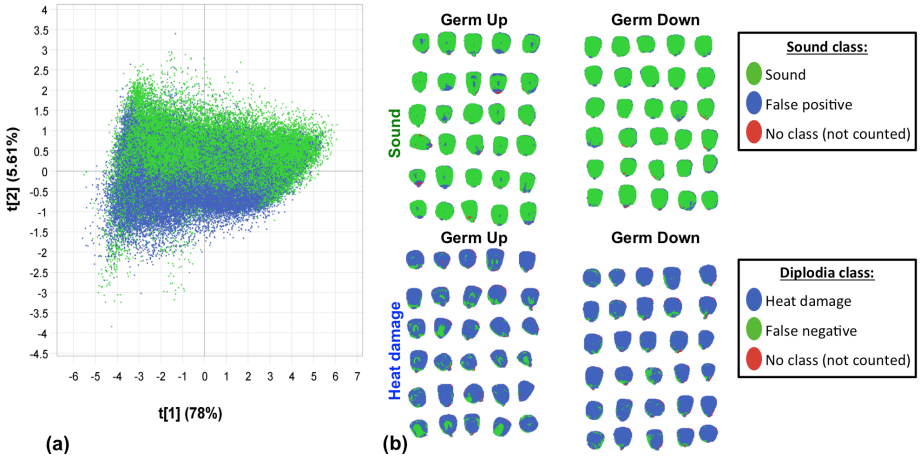

Figure 3.18 Pixel-wise PLS-DA classification resulted in well classified heat damage class vs. sound class (90.72% classification accuracy). Fair separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (78% SS) vs. 2 (6% SS); (b) Classification image.

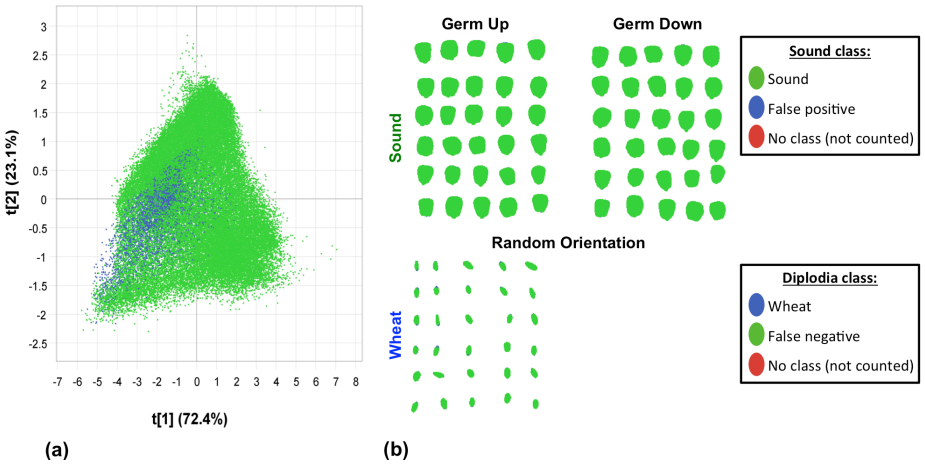

Figure 3.19 Pixel-wise PLS-DA classification resulted in extremely poorly classified wheat class vs. sound class (90.87% classification accuracy; 2.38% sensitivity). Minimal separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (72% SS) vs. 2 (23% SS); (b) Classification image.

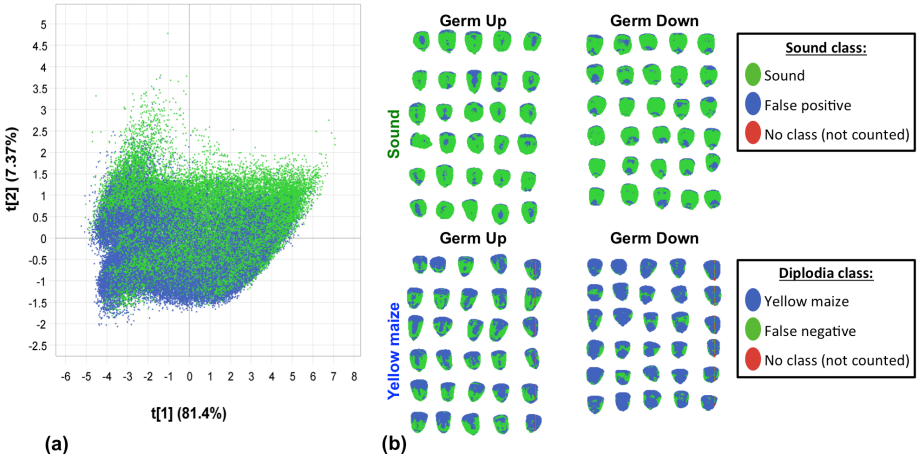

Figure 3.20 Pixel-wise PLS-DA classification resulted in poorly classified yellow maize class vs. sound class (75.32% classification accuracy). Minimal separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (81% SS) vs. 2 (7% SS); (b) Classification image.

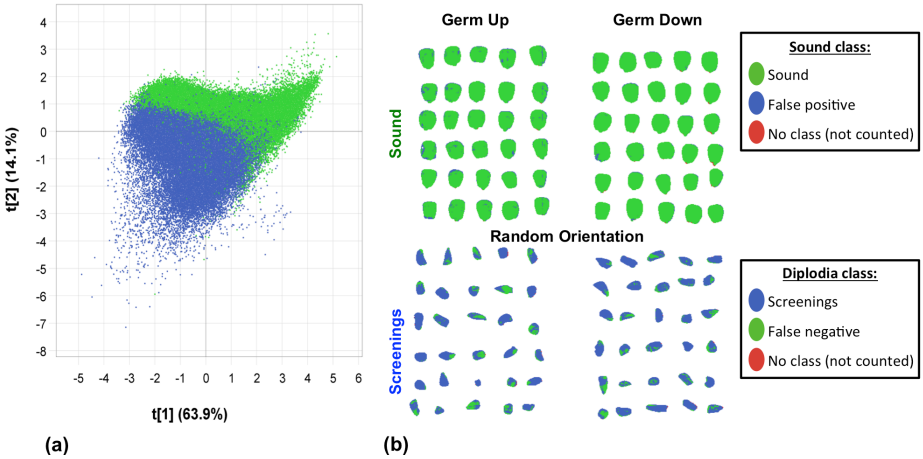

Figure 3.21 Pixel-wise PLS-DA classification resulted in well classified screenings class vs. sound class (91.58% classification accuracy). Fair separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (64% SS) vs. 2 (14% SS); (b) Classification image.

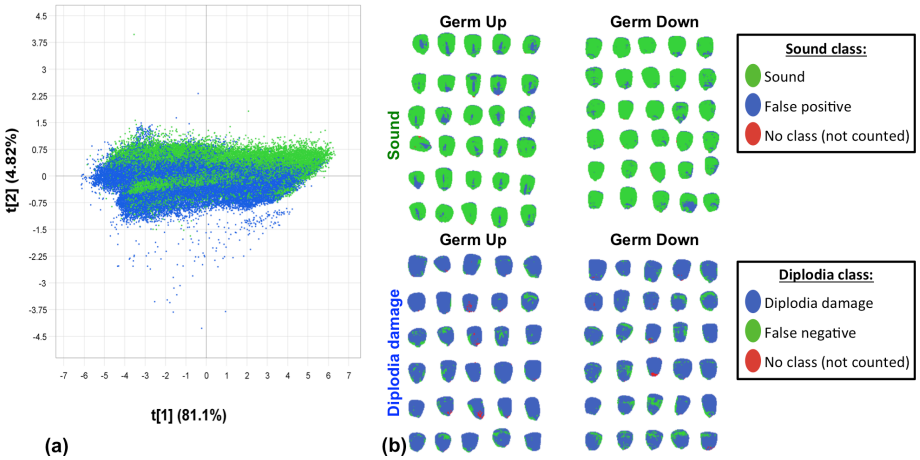

Figure 3.22 Pixel-wise PLS-DA classification resulted in well classified *Diplodia* damage class vs. sound class (85.22% classification accuracy). Slight separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (81% SS) vs. 2 (5% SS); (b) Classification image.

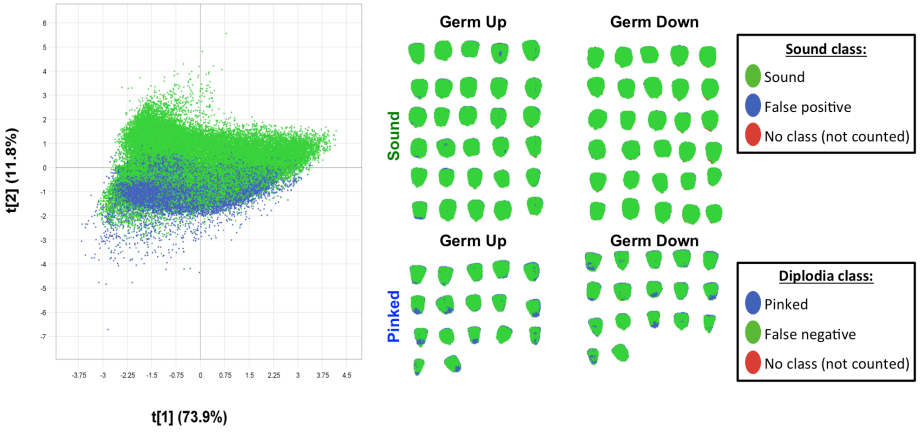

Figure 3.23 Pixel-wise PLS-DA classification resulted in extremely poorly classified pinked class vs. sound class (62.94% classification accuracy). Slight separation of classes was observed. (a) PLS-DA score plot of PLS factor 1 (74% SS) vs. 2 (12% SS); (b) Classification image.

The pixel-wise PLS-DA score plot visualises the projection of the validation image's pixels according to the calibrated model. The pixel-wise PLS-DA score plots for PLS factors 1 vs. 2 for heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are given in Fig. 3.18a to 3.23a. The score plots were indicative of the limited classification abilities of the models. Although the plots exhibit slightly clearer separations of class's pixels than in the pixel-wise PCA results, there was still an overlap between the majority of points. Some separation was apparent, notably in the screenings (Fig. 3.21a) and *Diplodia* damage classes (Fig. 3.22a). In general, the separation was observed to occur in the directions of both PLS factors 1 and 2, where the sound class was associated with the positive scores of both factors and the undesirable material class with the negative scores.

The considerable overlap presented throughout the PLS-DA score plots was accompanied by a large number of pixels being incorrectly classified, as seen in the pixel-wise classification images (Fig. 3.18-3.23b). The pixel-wise classification images for heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are given in Fig. 3.18b to 3.23b. As the pixel-wise classification images comprised small pixels, it was impossible to compare the results within the germ-up and -down image pairs. For these analyses, the results of correct and incorrect predictions were simply averaged for the whole image (all kernels). Although a large number of errors occurred, the majority of pixels were generally correctly predicted. The errors tended to arise in regions of the kernel that were very similar in both classes. This may be due to the defect not invading these particular regions, or simply because differences between the classes were not distinct enough to overcome the substantial variation within a kernel's components. This may be observed in the pixel-wise classification image of heat damage (Fig. 3.18b), where the pixels on the prominent germ (germ-up image) were widely misclassified. Similarly, certain portions of the endosperm in the sound maize and yellow maize were easily confused (Fig. 3.20b). The classification model seems to have been calibrated on the basis of hard and soft endosperm, where

Table 3.2 Results of pixel-wise PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes

Undesirable material	R^2	PLS factors	Classification accuracy (%)	False negatives (%)	False positives (%)	Sensitivity (%)	Specificity (%)
Fusarium damage	0.534	5	82.67	9.35	7.97	84.38	86.07
Diplodia damage	0.337	5	85.22	7.83	6.95	80.04	84.99
Water damage	0.309	4	78.78	13.17	8.05	92.45	89.19
Rodent damage	0.502	3	82.95	6.13	10.92	74.64	83.26
Heat damage	0.513	4	90.72	3.54	5.74	71.99	86.02
Screenings	0.589	3	91.58	2.85	5.57	88.77	92.54
Pinked maize	0.450	4	62.94	14.23	22.83	40.91	69.93
Yellow maize	0.239	4	75.32	9.02	15.65	90.17	96.20
Plant material	0.856	2	86.86	6.43	6.71	98.49	99.63
Wheat	0.207	4	90.87	2.09	7.04	2.38	92.80
Sorghum	0.388	5	96.04	0.26	3.70	68.68	92.80
Soy	0.795	3	99.46	0.23	0.31	83.17	89.15
Sunflower	0.944	2	99.65	0.00	0.35	77.85	73.59

R^2: Coefficient of determination

PCs: Principal components

the outer hard endosperm was associated strongly with the sound class (green) and the inner soft endosperm was associated with the yellow maize class (blue). The classification of wheat was the poorest outcome of all analyses (Fig. 3.19b). The classification accuracy was not as severely affected (90.87%) as the much smaller wheat kernels contributed very few pixels toward the averaged value relative to the large maize kernels. However, the incredibly low sensitivity of the model (2.38%) reveals how poorly suited the model was for the prediction of wheat content in a maize sample.

The low R^2-values of the pixel-wise PLS-DA models (Table 3.2) demonstrated the difficulty of differentiating between classes when there was a large overlap of similar chemical composition. The number of PLS factors required for optimal calibration of the models varied from PLS factor 2 to 5. The optimal number of PLS factors was determined by evaluating the % SS and subsequent changes in R^2 value. Unexpectedly, some models did not exhibit improved fit beyond the PLS factors 1 and 2. Rodent damage, which displays a large overlap of chemical characteristics with the sound class, was calibrated using 3 PLS factors as the further factors resulted in reduced fit. On the other hand, some foreign matter classes that were not closely related to the sound class, such as sorghum, exhibited best fit with PLS factor 5. The only classes with notably high R^2-values were sunflower seeds (0.944) and plant material (0.856), most likely due to the complete difference in chemical composition.

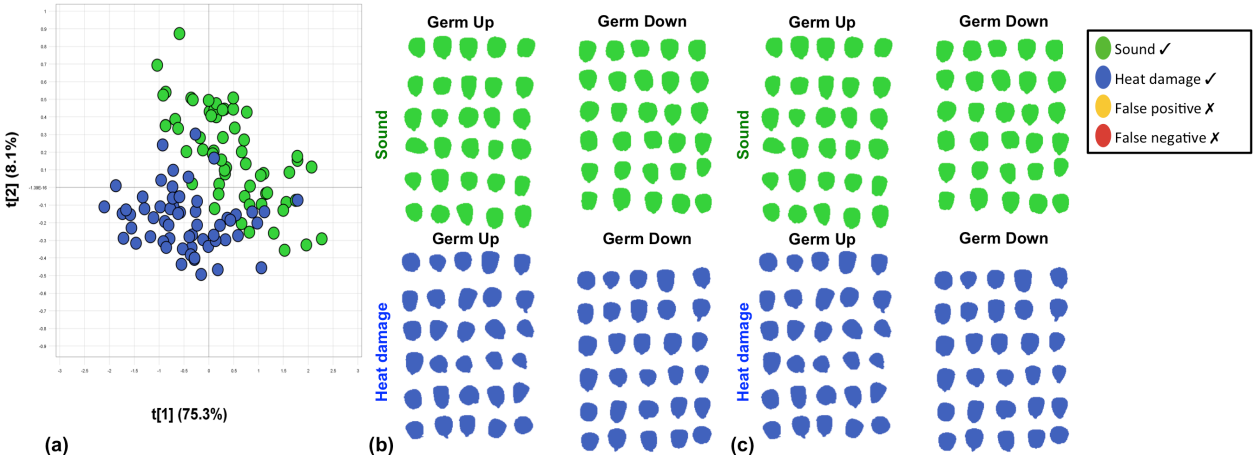

Figure 3.24 Object-wise PLS-DA classification resulted in perfectly classified heat damage class vs. sound class (100% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (75% SS) vs. 2 (8% SS); (b) Unprocessed classification image; and (c) Overall classification image.

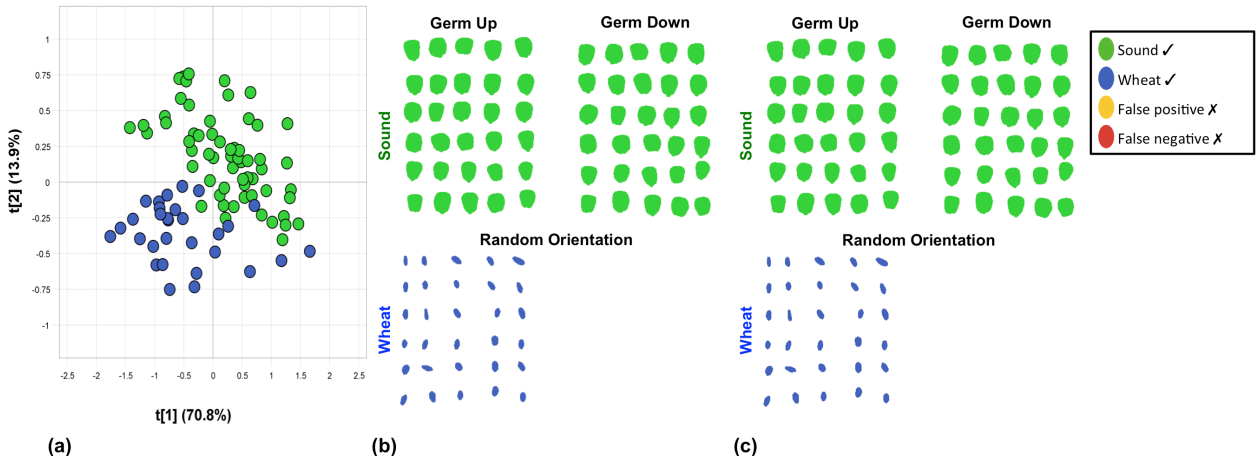

Figure 3.25 Object-wise PLS-DA classification resulted in perfectly classified wheat class vs. sound class (100% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (71% SS) vs. 2 (14% SS); (b) Unprocessed classification image; and (c) Overall classification image.

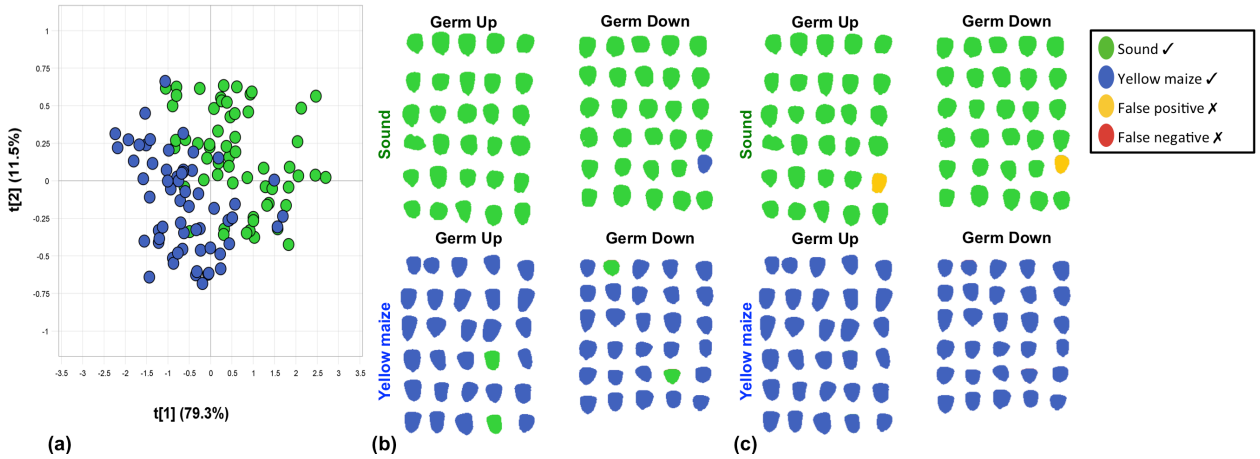

Figure 3.26 Object-wise PLS-DA classification resulted in well classified yellow maize class vs. sound class (98.33% classification accuracy). (a) PLS-DA score plot PLS factor 1 (79% SS) vs. 2 (12% SS); (b) Unprocessed classification image; and (c) Overall classification image.

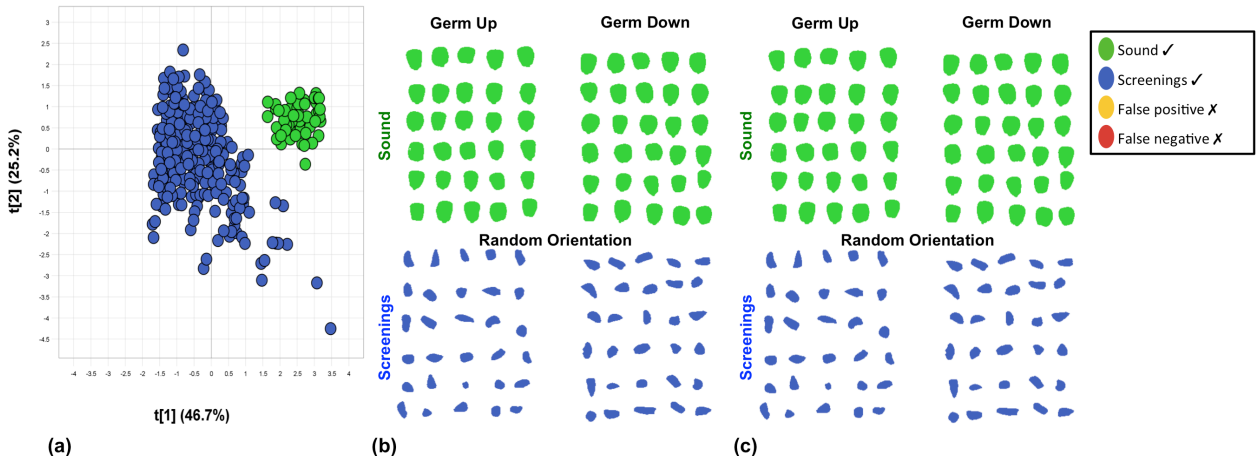

Figure 3.27 Object-wise PLS-DA classification resulted in perfectly classified screenings class vs. sound class (100% classification accuracy). (a) PLS-DA score plot PLS factor 1 (47% SS) vs. 2 (25% SS); (b) Unprocessed classification image; and (c) Overall classification image.

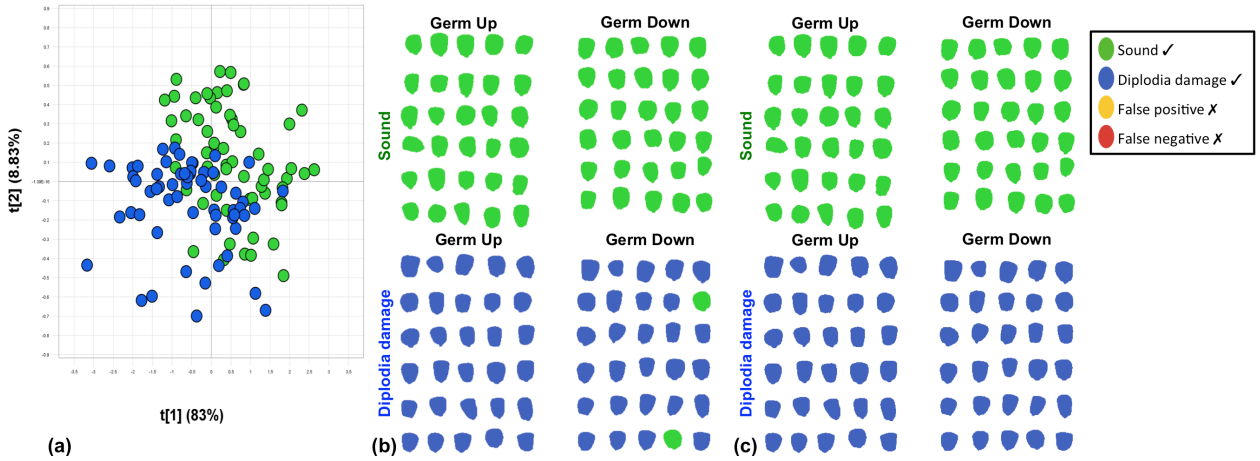

Figure 3.28 Object-wise PLS-DA classification resulted in perfectly classified *Diplodia* damage class vs. sound class (100% classification accuracy). (a) PLS-DA score plot PLS factor 1 (83% SS) vs. 2 (9% SS); (b) Unprocessed classification image; and (c) Overall classification image.

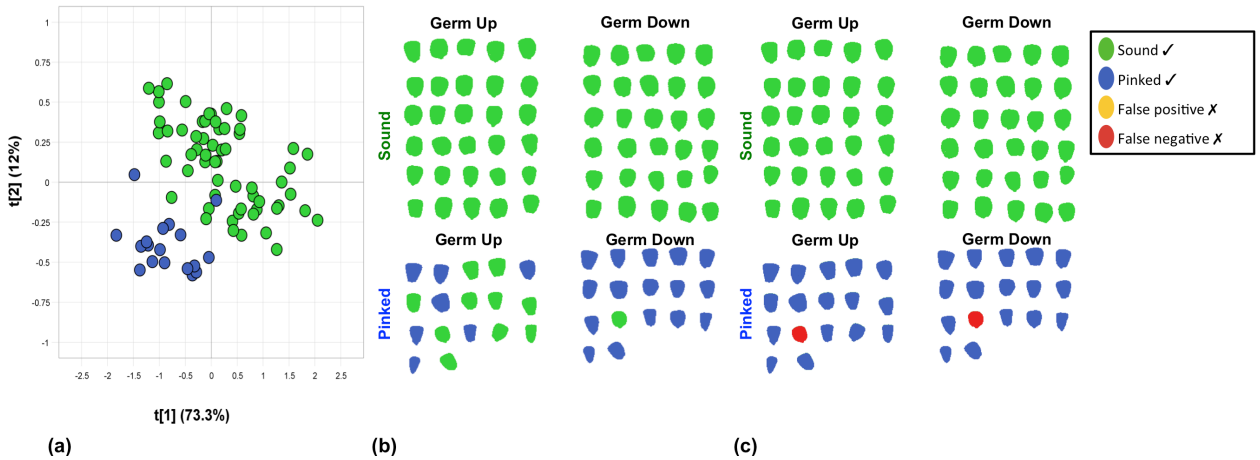

Figure 3.29 Object-wise PLS-DA classification resulted in well classified pinked class vs. sound class (98.15% classification accuracy). (a) PLS-DA score plot of PLS factor 1 (73% SS) vs. 2 (12% SS); (b) Unprocessed classification image; and (c) Overall classification image.

The object-wise PLS-DA score plots for heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are given in Fig. 3.24a to 3.29a. The score plots demonstrate very little overlap between the classes, indicating excellent model calibration. The score plot of the screenings class (Fig. 3.27a) exhibited the most striking separation, with two tightly arranged clusters on either side of the PLS factor 1 axis. There was little separation in the direction of PLS factor 2. Similarly, the foreign matter classes (wheat (Fig. 3.25a), soy, sorghum, sunflower and plant material) and yellow maize (Fig. 3.26a) exhibited the majority of separation in the direction of PLS factor 1, while the spread across PLS factor 2 was generally intra-class separation. The remaining white maize defect classes, demonstrated by heat damage (Fig. 3.24a), *Diplodia* damage (Fig. 3.28a) and pinked kernels (Fig. 3.29a), tended to show separation occurring in the direction of both PLS factor 1 and 2. The sound class was associated with the positive scores in both factors and the defect class with the negative scores.

Table 3.3 Results of object-wise PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes

Undesirable material	R^2	Q^2	PLS factors	Classification accuracy (%)	False negatives (%)	False positives (%)	Sensitivity (%)	Specificity (%)
Fusarium damage	0.911	0.872	5	98.33	0.00	1.67	100.00	96.67
Diplodia damage	0.836	0.814	5	100.00	0.00	0.00	100.00	100.00
Water damage	0.780	0.754	4	98.33	1.67	0.00	96.67	100.00
Rodent damage	0.935	0.926	3	98.33	1.67	0.00	96.67	100.00
Heat damage	0.886	0.870	4	100.00	0.00	0.00	100.00	100.00
Screenings	0.878	0.855	3	100.00	0.00	0.00	100.00	100.00
Pinked maize	0.934	0.917	4	98.15	1.85	0.00	95.83	100.00
Yellow maize	0.810	0.786	4	98.33	0.00	1.67	100.00	96.67
Plant material	0.956	0.938	2	100.00	0.00	0.00	100.00	100.00
Wheat	0.921	0.905	4	100.00	0.00	0.00	100.00	100.00
Sorghum	0.951	0.936	5	100.00	0.00	0.00	100.00	100.00
Soy	0.974	0.966	3	100.00	0.00	0.00	100.00	100.00
Sunflower	0.995	0.994	2	100.00	0.00	0.00	100.00	100.00

R^2: Coefficient of determination

Q^2: Cross-validated coefficient of determination

PCs: Principal components

The object-wise PLS-DA analyses yielded promising results and clearly demonstrated the advantages of using an object as the lowest unit of measurement. The majority of classes were predicted perfectly, and no more than one error per two-way classification occurred in the remaining analyses. The object-wise unaltered and overall classification images of heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are given in Fig. 3.24b to 3.29b and Fig. 3.24c to 3.29c, respectively. The unaltered classification images display the classification of each object, where green classifications are sound class and blue classifications are the respective undesirable material class. All objects in the green (sound) half should be coloured green, while all objects in the blue (defect) half should be coloured blue. If objects appear as the incorrect colour, it was determined if a false positive (yellow) or false negative (red) occurred, according to Table 3.3. Classifications were totalled as follows: both sides classified 'sound' was counted as 'sound'; both sides classified 'defect' was counted as 'defect'; and one side classified 'sound' and one side classified 'defect' was counted as 'defect'. As each kernel received one overall classification, the germ-up and -down classification was identical. Thus, only one section (germ-up or -down) was requisite for totalling. The classification images demonstrate the excellent outcome of the classification process. Most analyses show perfect classification, where all sound kernels were predicted as sound (green) and all undesirable materials were predicted as defective (blue). The pinked class (Fig. 3.29) demonstrated the importance of capturing images of the kernels in both the germ-up and -down orientation. In the germ-up image half, 10 of the 24 pinked kernels were misclassified, likely due to the defect being less severe on the germ-up side of the kernel. However, when the kernels were turned over, 9 of the 10 previously misclassified kernels were correctly identified as defective. Thus, only one false negative occurred. Similarly, two kernels in the *Diplodia* class (Fig. 3.28) were misclassified when facing germ-up but were both correctly identified on the other side, while the yellow maize class (Fig. 3.26) exhibited misclassification of 4 different kernels (2 on each side) but overall all kernels were correctly identified on at least one side. In the six analyses used as examples, one false positive occurred in the analysis of yellow maize vs. sound maize (Fig. 3.26). This demonstrated that due to the experimental set-up, once a sound kernel was misclassified on either side, it was immediately deemed incorrectly classified (false positive). The sensitivity and specificity of all of the perfectly classified classes was 100%, while the remaining classes (*Fusarium* damage, water damage, rodent damage, pinked maize and yellow maize) had reduced specificity and sensitivity due to their classification errors.

The R^2 and Q^2 values were generally quite high, ranging 0.780 to 0.995 and 0.754 to 0.994, respectively. The R^2 value of each class was considerably higher in the object-wise approach than the pixel-wise approach. This reaffirms that an object-wise approach was able to cut through the noise of class overlap and identify meaningful differences between two classes for the purpose of classification.

CONCLUSION

Hyperspectral imaging was successfully used to separate sound white maize from the 13 major classes of undesirable materials encountered during industry grading practices. Two approaches to the data analysis, namely pixel-wise and object-wise based analyses, were investigated. The pixel-wise PCA results were influenced by large heterogeneity of the kernels, which overshadowed the less prominent differences in chemical composition between the closely related classes. The main sources of variance identified in the initial PCs were due to differences within the kernels (the various anatomical kernel components) and not due to differences between the classes. The object-wise approach eliminated any possibility of identifying variance within a kernel by using the whole kernel as the smallest unit of measurement. By calculating an average spectrum per kernel, the object-wise PCA analysis achieved moderate separation of the classes. PLS-DA classification models were developed for two-way classification of sound maize vs. each undesirable material. The pixel-wise PLS-DA analyses yielded classifications that were not sufficiently accurate for industry implementation. The coefficients of classification (R^2) were low (0.207–0.944) and a large number of errors occurred with the classification accuracy ranging 62.94–99.65%. The object-wise PLS-DA analyses were successful in achieving accurate separations. Of the 13 analyses, 8 achieved perfect results, and the remaining 5 analyses contained only 1 error each per analysis of *ca.* 60 kernels. As the calibration and validation of the models were conducted using image data from independent samples, the results have demonstrated the accuracy of the model calibrations. Hyperspectral imaging paired with object-wise multivariate data analysis was shown to be highly suitable for separating sound white maize from common undesirable materials, and has a promising future in cereal grading applications. One drawback of hyperspectral image analysis is the high cost of the instrument and longer image acquisition and processing times. In order to minimise these limiting factors, an instrument utilising a limited number of key wavelengths, as opposed to the full range of the current instrument, could be used. Investigation into white maize grading using a multispectral imaging system is thus recommended for industry use. Hyperspectral imaging, however, was found to provide a large and rich dataset that was ideally suited for the purpose of this laboratory-scale research study.

REFERENCES

Amigo, J. M., Martí, I. & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality. In: *Data Handling in Science and Technology.* Pp. 343-370. Amsterdam, The Netherlands: Elsevier Science.

Cogdill, R. P., Hurburgh, C. R., Rippke, G. R., Bajic, S. J., Jones, R. W., McClelland, J. F., Jensen, T. C. & Liu, J. (2004). Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE*, **47**, 311.

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, **144**, 64-71.

Delwiche, S. R. & Hareland, G. A. (2004). Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry*, **81**, 643-649.

Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Elmasry, G. & Sun, D. W. (2010). Principles of hyperspectral imaging technology. In: *Hyperspectral imaging for food quality analysis and control* (edited by D.W Sun.). Pp. 3-44. London, UK: Elsevier.

Esbensen, K. & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, **7**, 67-86.

Esbensen, K. H., Guyot, D., Westad, F. & Houmoller, L. P. (2002). Principal Component Analysis (PCA) - Introduction. In: *Multivariate data analysis - In practice: An introduction to multivariate data analysis and experimental design*. Pp. 19-74. Camo ASA.

Fernández-Ibañez, V., Soldado, A., Martínez-Fernández, A. & De la Roza-Delgado, B. (2009). Application of near infrared spectroscopy for rapid detection of aflatoxin B1 in maize and barley as analytical quality assessment. *Food chemistry*, **113**, 629-634.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G. & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, **18**, 590-598.

Johnson, L. A. (2000). The Major Cereal of the Americas. In: *Handbook of Cereal Science and Technology, Revised and Expanded* (edited by K. Kulp & J. G. Ponte). Pp. 31-80. New York, USA: CRC Press.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126-135.

Mahesh, S., Jayas, D. S., Paliwal, J. & White, N. D. G. (2015). Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat. *Food and Bioprocess Technology*, **8**, 31-40.

Manley, M., Du Toit, G. & Geladi, P. (2011). Tracking diffusion of conditioning water in single wheat kernels of different hardnesses by near infrared hyperspectral imaging. *Analytica Chimica Acta*, **686**, 64-75.

Manley, M., Williams, P., Nilsson, D. & Geladi, P. (2009). Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (Zea maize L.) kernels. *Journal of Agricultural and Food Chemistry*, **57**, 8761-8769.

Mark, H. & Workman Jr, J. (2003). Derivatives in Spectroscopy Part I: The Behavior of the Derivative. *Spectroscopy Eugene*, **18**, 32-37.

McGoverin, C. & Manley, M. (2012). Classification of maize kernel hardness using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **20**, 529.

McGoverin, C. M., Engelbrecht, P., Geladi, P. & Manley, M. (2011). Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. *Analytical and Bioanalytical Chemistry*, **401**, 2283-2289.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). Data compression by PCR and PLS. In: *A user-friendly guide to multivariate calibration and classification*. Pp. 27-38. Chichester, U.K.: NIR Publications.

Park, B. & Lu, R. (2015). Introduction (Image and Spectral Analysis Techniques). In: *Hyperspectral imaging technology in food and agriculture* (edited by B. Park & R. Lu). Pp. 3-8. New York, USA: Springer.

Rinnan, Å., van den Berg, F. & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, **28**, 1201-1222.

Serna-Saldivar, S. O. (2010). Physical Properties, Grading, and Speciality Grains. In*: Cereal grains: properties, processing, and nutritional attributes*. Pp. 43-80. Boca Raton, USA: CRC Press.

Singh, C. B., Jayas, D. S., Paliwal, J. & White, N. D. G. (2010). Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Computers and Electronics in Agriculture*, **7**3, 118-125.

Tsai, F. & Philpot, W. (1998). Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, **66**, 41-51.

Wang, L., Liu, D., Pu, H., Sun, D. W., Gao, W. & Xiong, Z. (2014). Use of Hyperspectral Imaging to Discriminate the Variety and Quality of Rice. *Food Analytical Methods*, **8**, 515-523.

Wang, L., Sun, D. W., Pu, H. & Zhu, Z. (2015). Application of Hyperspectral Imaging to Discriminate the Variety of Maize Seeds. *Food Analytical Methods*, **9**, 1-10.

Weinstock, B. A., Janni, J., Hagen, L. & Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (*Zea mays L.)* kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*, **60**, 9-16.

Williams, P., Geladi, P., Fox, G. & Manley, M. (2009). Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Analytica Chimica Acta*, **653**, 121-130.

Williams, P. J., Geladi, P., Britz, T. J. & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, **55**, 272-278.

CHAPTER 4

CHARACTERISATION OF WHITE MAIZE KERNELS USING MULTISPECTRAL IMAGING

ABSTRACT

Multispectral imaging with object-wise multivariate image analysis was evaluated for its potential to grade whole white maize kernels. Although hyperspectral imaging has been shown to be capable of separating sound maize and undesirable materials, the technique was not rapid enough and potentially too expensive for on-line industry application. Thus, the aim was to achieve similar classification capabilities with a simpler and less expensive spectral imaging system. The types of defective materials regarded were divided into 13 classes, and were imaged with a multispectral imaging instrument spanning the UV, visible and NIR regions (19 spectral points ranging 375–970 nm). Two-way classifications were performed to distinguish various types of defective kernels, pinked kernels, yellow maize and foreign matter from sound white maize. Object-wise partial least squares discriminant analysis (PLS-DA) models were developed and validated with an independent image data set for 13 undesirable material classes vs. sound white maize. Results demonstrated good performance in distinguishing between the sound maize undesirable materials, with coefficients of determination (R^2) and classification accuracies ranging from 0.46 to 0.95 and 83 to 100%, respectively. The closely related 6 white maize defects and pinked white maize were more challenging to separate from sound white maize, and some errors occurred during these analyses. The yellow maize and 5 foreign matter classes were easily separated from sound maize with 100% accuracy. The wavebands related to green, yellow and orange colour (505, 525, 570 and 590 nm), and the NIR wavebands 890 and 940 nm (associated with fat) and 970 nm (associated with water) were generally identified as important features throughout the classes. The classification accuracy achieved with this instrument was generally lower than with the hyperspectral instrument, but the cost and scan time was substantially reduced.

INTRODUCTION

Multispectral imaging is a form of spectral imaging in which a relatively small number of wavebands are used. As with hyperspectral imaging, it is a non-destructive technique that combines spectroscopy and computer imaging, enabling chemical and spatial information to be acquired simultaneously (Ariana *et al.*, 2006; Dissing *et al.*, 2009). By utilising only a small number of key wavebands rather than a continuous range, multispectral instruments are simple, rapid and inexpensive. Current hyperspectral instruments have a scan time of approximately 8 seconds, which is too slow for the online industry implementation (Ariana *et al.*, 2006) where thousands of samples need to be assessed per minute. Since multispectral imaging is based on far fewer spectral bands, faster image acquisition can be expected. Furthermore, the inclusion of only a few discrete wavebands allows multispectral instruments to be manufactured at lower costs. If multispectral imaging can prove to be capable of similar applications as hyperspectral imaging at suitable accuracies, its higher throughput and lower cost is favourable for industry applications.

Multispectral systems show great promise for on-line cereal grading and evaluation (Liu *et al.*, 2014a). Although multispectral imaging has been less widely utilised, successful food applications include the evaluation of red meat (Ropodi *et al.*, 2013; Trinderup *et al.*, 2013; Ropodi *et al.*, 2015; Liu *et al.*, 2016a), fish (Ljungqvist *et al.*, 2012), prepared meal components (Dissing *et al.*, 2009), fruits and vegetables (Ariana *et al.*, 2006; Shrestha *et al.*, 2015; Liu *et al.*, 2014b), cereals (Liu *et al.*, 2014a; Liu *et al.*, 2016b; Bodevin *et al.*, 2009), and baked cereal products (Andresen *et al.*, 2013). Due to the ability to measure multiple components at the same time, an on-line multispectral system has the potential to replace multiple conventional chemical, microbial or physical tests with a single, automated image acquisition (Dissing *et al.*, 2009). The cereal industry would be able to move away from several current subjective and tedious manual classification and measuring methods, including grading.

Many applications of spectral imaging in cereal research utilise the near infrared (NIR) region, including the hyperspectral imaging study in Chapter 3. However, the visible or ultraviolet (UV) regions may also be used (Berman *et al.*, 2007; Del Fiore *et al.*, 2010; Wang *et al.*, 2015). The use of these regions is fitting, as the undesirable materials given in the South African grading legislation are all identifiable using human visual inspection. While chemical differences may be used to differentiate classes, as is done in NIR spectral imaging, visible differences will certainly be apparent between the samples. An instrument using the visible region could take advantage of these visible differences. This multispectral imaging study did not utilise the NIR region alone, but incorporated UV and visible wavebands. An advantage of using visible wavebands instead of the NIR is a reduction in cost of the camera, which could be highly beneficial to the industry applications of the technology.

In the previous research chapter (Chapter 3) the use of object-wise image analysis was successfully explored. Pixels from the same objects (e.g. kernels) were grouped together and

spectral features for the whole object are calculated (Kucheryavskiy, 2013). This method has proved more effective in cases where objects from different classes have many similar pixels (Kucheryavskiy, 2013; Williams and Kucheryavskiy, 2016). These similar pixels will be associated with both classes when a classification model is calibrated. A pixel-wise classification model based on this data was found to be unstable and result in poor classifications. Instead, by predicting objects, calibrations are instead based on patterns the objects' pixels share.

The aim of this study was to investigate if multispectral imaging with object-wise PLS-DA can offer comparable or improved classifications in comparison with hyperspectral imaging (Chapter 3). Whole sound white maize kernels were separated from common undesirable materials encountered in the South African maize industry.

MATERIALS AND METHODS

Samples

Maize kernels and undesirable materials were obtained from the South African Grain Laboratory (SAGL, Pretoria, South Africa) and were graded visually by expert graders at the SAGL according to South African grading regulations (Department of Agriculture, 2009). The 13 most prolific undesirable materials encountered in the South African maize industry were selected as classes to be distinguished from the sound maize class (see Fig. 3.1 in previous chapter). Classes included defective white maize (heat damage, water damage, screenings/broken kernels, *Fusarium* fungal damage and *Diplodia* fungal damage); pinked white maize; other colour maize (i.e. yellow maize); and foreign materials (wheat, soy, sunflower seeds, sorghum and maize plant material). The classes relate to the South African grading legislation, which outlines 4 main categories. Either the main category or several sub-categories were used as classes.

Multispectral system

Multispectral images were acquired on a VideometerLab2 (Videometer, Hørsholm, Denmark) multispectral imaging system. The samples were placed beneath an integrating or Ulbricht sphere, with a camera located in the top of the sphere. During image capture, the sphere closes over the sample stage to create optically closed conditions, allowing even lighting with minimal shadows and specular reflection. Samples were illuminated by 19 high power light emitting diodes (LEDs) that were evenly spaced along the equator of the sphere. A spectral range of 375 to 970 nm was utilised, at specific wavelengths, as outlined in Table 4.1. The LEDs strobe successively, resulting in an image for each LED. The images consisted of 2056 x 2056 pixels, with a high spatial resolution of about 45 µm/pixel. Instrument calibration was performed approximately once an hour using two reflectance targets (25 and 75% reflectance) and one geometric target to obtain the optimal dynamic range for each LED and minimize lens distortions.

Table 4.1 The wavebands utilised by the VideometerLab2 instrument and the associated region in the electromagnetic spectrum.

Variable	Waveband (nm)	Region	Variable	Waveband (nm)	Region
Variable 1	375	UV	Variable 11	645	Visible (Red)
Variable 2	405	Visible (Violet)	Variable 12	660	Visible (Red)
Variable 3	435	Visible (Indigo)	Variable 13	700	Visible (Red)
Variable 4	450	Visible (Blue)	Variable 14	780	NIR
Variable 5	470	Visible (Blue)	Variable 15	850	NIR
Variable 6	505	Visible (Green)	Variable 16	870	NIR
Variable 7	525	Visible (Green)	Variable 17	890	NIR
Variable 8	570	Visible (Yellow)	Variable 18	940	NIR
Variable 9	590	Visible (Orange)	Variable 19	970	NIR
Variable 10	630	Visible (Red)			

Image acquisition

For the calibration set, images of each of the 13 undesirable material classes and the sound maize kernels were captured. These data sets consisted of 17 sound maize kernels (top 3-4 rows in the image) and 18 undesirable material kernels/objects of a single class (bottom 3-4 rows in the image), giving a total of 35 objects per image. The only exception was the screenings class, which consisted of 30 objects instead of 18, due to their small sizes. For the classes of maize kernels (sound maize; *Fusarium* damage; *Diplodia* damage; water damage; heat damage; rodent damage; and yellow maize), the kernels were first imaged with the germ facing upwards (towards the camera), and a second time with the germ facing downwards (away from the camera). Thus the germ-up and -down oriented versions of the images consisted of the same kernels in the same positions. For the foreign matter classes (plant material, wheat, sorghum, soy and sunflower seeds), only the sound maize kernels were orientated as before, but completely different undesirable material objects were used between versions, as the foreign matter often had no obvious germ. The validation set was acquired in the same manner on new sets of kernels and objects for each of the 13 undesirable material classes.

Multispectral image analysis

Images were analysed using the Evince v.2.7.0 (UmBio AB, Umeå, Sweden) spectral image analysis software package. The germ-up and -down images for each pair (sound kernels and undesirable materials) were mosaiced and each image mosaic was analysed individually. A summarised illustration of the image analysis process is given in Fig. 4.2.

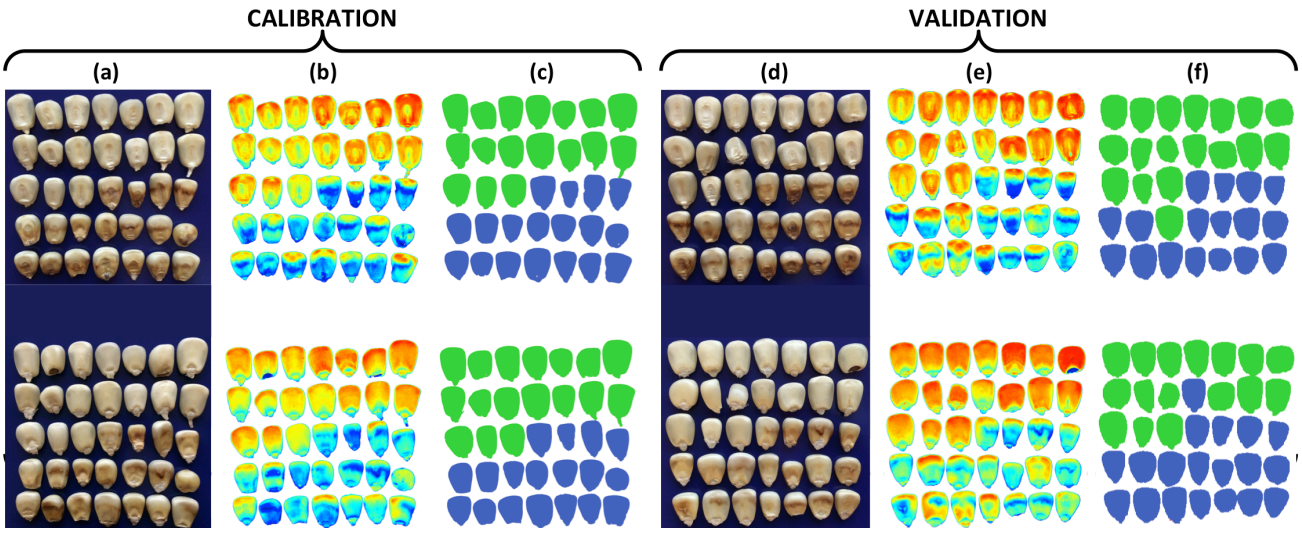

Figure 4.1 Summary of image analysis, illustrated using sound class (first 17 kernels) vs. yellow class (second 18 kernels). The top half of each section (a–f) is germ-up, and the bottom half is germ-down. Each half contains the same kernels in the same positions, to allow for inspection of both sides. (a) Digital image of calibration samples arranged in known order of classes; (b) PCA score image of calibration samples after pre-processing; (c) Samples were assigned classes (sound as green and defect as blue) and PLS-DA model was calibrated; (d) Digital image of independent validation samples arranged in known order of classes; (e) PCA score image of validation samples after pre-processing; and (f) PLS-DA model applied to generate classification image.

Image correction and cleaning

The image calibration and correction from reflectance to pseudo-absorbance was done automatically in the Evince software package as in Chapter 3. PCA was applied to the mean-centred absorbance mosaic images. Score plots and score images were used interactively to identify unwanted pixels, e.g. outliers, sample stage background, dead pixels, shading errors and edge effects, which were subsequently removed from the dataset (Esbensen and Geladi, 1989). PCA was recalculated with additional components after obtaining a cleaned image. The cleaned image was used in subsequent analysis.

Principal component analysis

After various pre-treatments were evaluated, standard normal variate (SNV) transformation was chosen and applied to the data. Data were analysed with pixel-wise and object-wise PCA, and the PC score images, score plots and loading line plots were examined.

Partial least squares discriminant analysis

PLS-DA models were calculated to distinguish between the two classes, namely sound and respective undesirable material. The aim of PLS-DA is to model the relationship between a set of predictor variables (**X**), namely the spectral data, and a set of responses (**Y**) (Næs *et al.*, 2002). The response was the assignment of an object to a mutually exclusive class, sound or undesirable

material. Object-wise PLS-DA models were calculated with full cross-validation. The mean spectrum of all pixels in an object was used during object-wise analyses. The PLS-DA models were applied to the validation image data, and a classification image was generated, referred to as the unaltered classification image in later sections.

Correct and incorrect predictions, with reference to the kernel's assigned classes, were totalled for each PLS-DA classification image, as in the object-wise results for Chapter 3 and illustrated in Fig. 4.3. If the overall classification was correct, it was counted as a correct classification, forming part of the correct classification accuracy (%). If it was incorrect, it was classified as either a false positive or false negative, forming part of the false positive error (%) or false negative error (%), respectively.

Classification accuracy, false positive error, false negative error, sensitivity and specificity were calculated as in Chapter 3. An undesirable material classification was a positive response and a sound classification was a negative response. A false positive occurred when a sound kernel was incorrectly classified as undesirable material, and a false negative was when an undesirable material was incorrectly classified sound. The sensitivity describes the probability that undesirable will be detected and correctly classified, and is sometimes described as the true positive rate. Specificity is the probability that a sound kernel will be classified correctly, also known as the true negative rate.

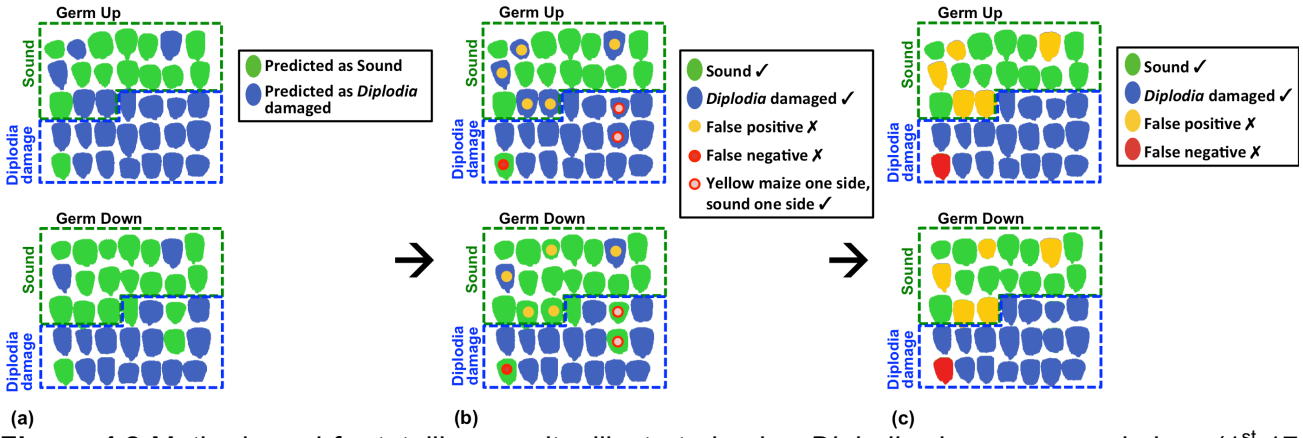

Figure 4.2 Method used for totalling results, illustrated using *Diplodia* class vs. sound class (1st 17 kernels). (a) Unprocessed classification image; (b) Totalling procedure for determining overall classification; and (c) Overall classification image.

RESULTS AND DISCUSSION

Multispectral image analysis

Principal component analysis

PCA was calculated to examine the qualitative difference between sound white maize and the various undesirable material classes. These were conducted on a pixel basis to facilitate brushing, as well as more in depth exploration. The classes heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are used to illustrate the PCA results, as given in Figs. 4.4 to 4.9. Note that the scores plots (a) are shaded according to density of plots, while the scores images (b) are shaded according to scores value where blue is negative and red is positive.

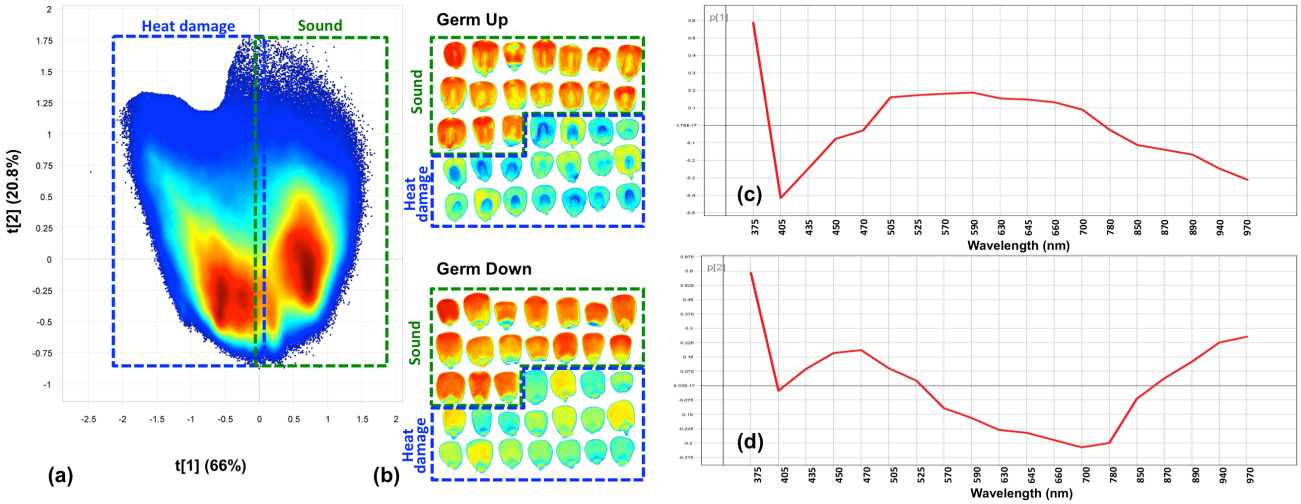

Figure 4.3 Well separated heat damage class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses.

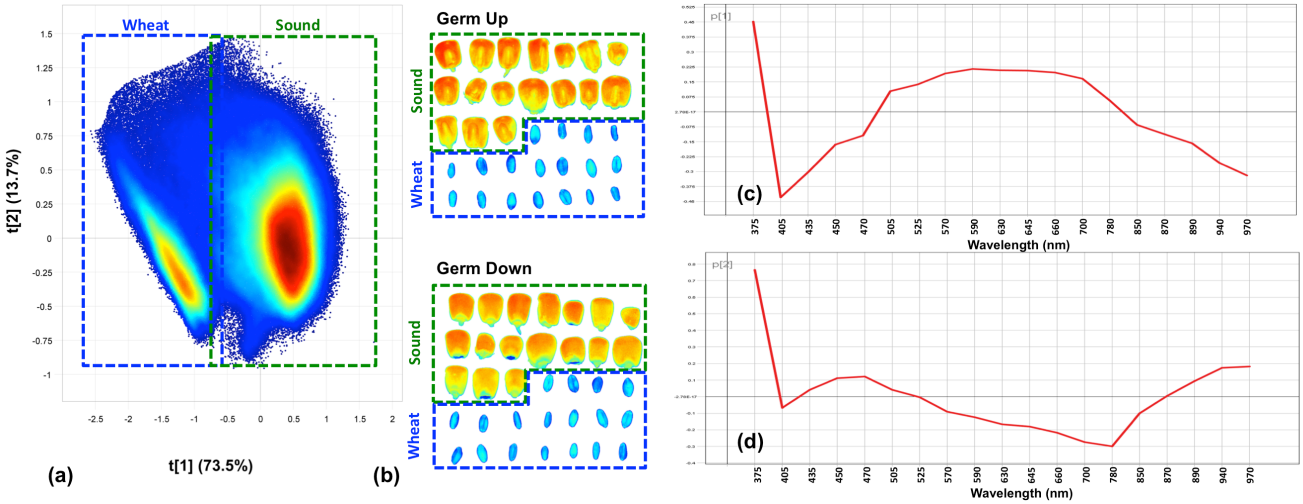

Figure 4.4 Well separated wheat class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses.

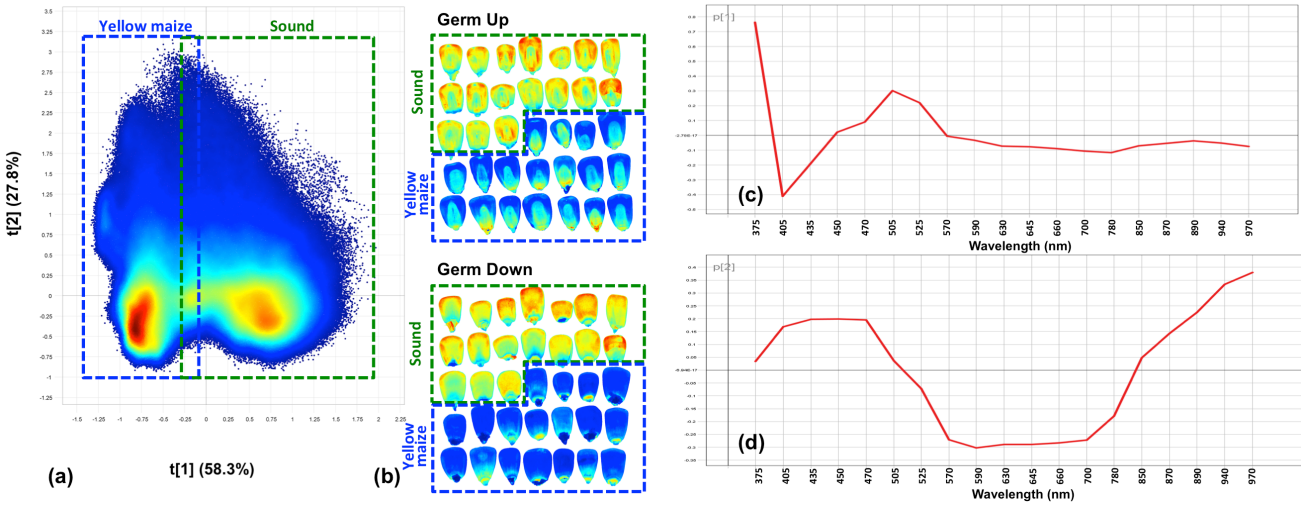

Figure 4.5 Well separated yellow maize class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 do not represent the general loadings line trend observed over most of the 13 PCA analyses.

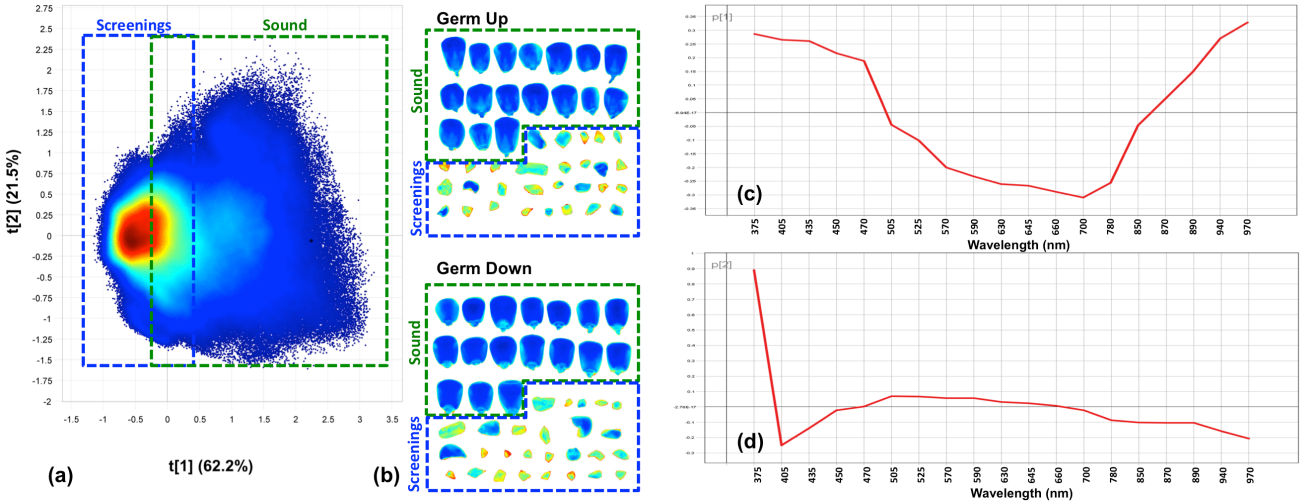

Figure 4.6 Moderately separated screenings class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the inverse of the general loadings line trend observed over most of the 13 PCA analyses.

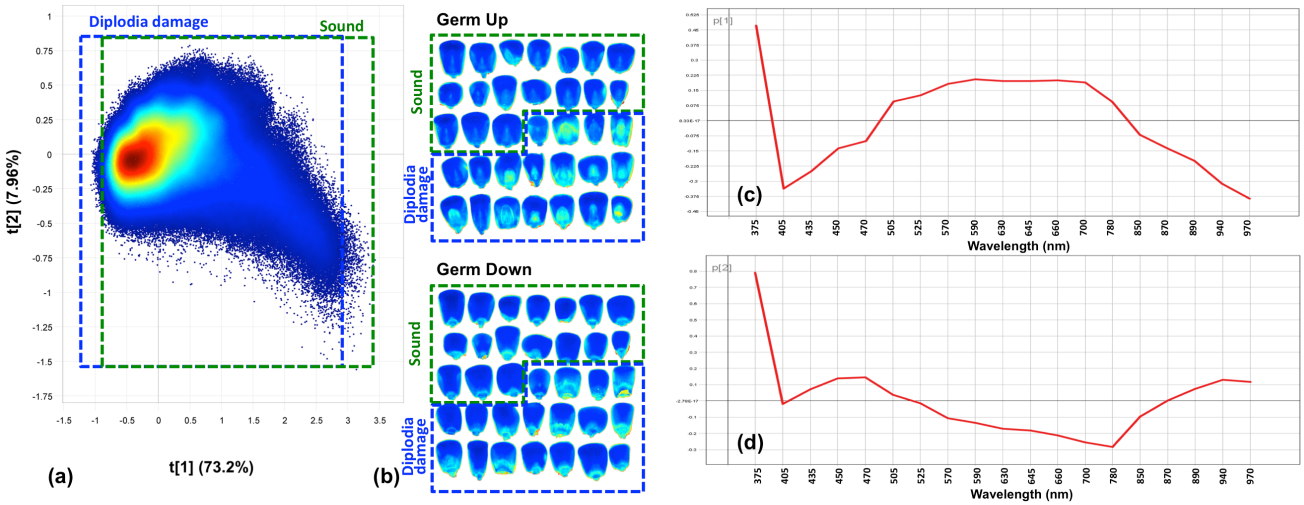

Figure 4.7 Not separated *Diplodia* damage class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 represent the general loadings line trend observed over most of the 13 PCA analyses.

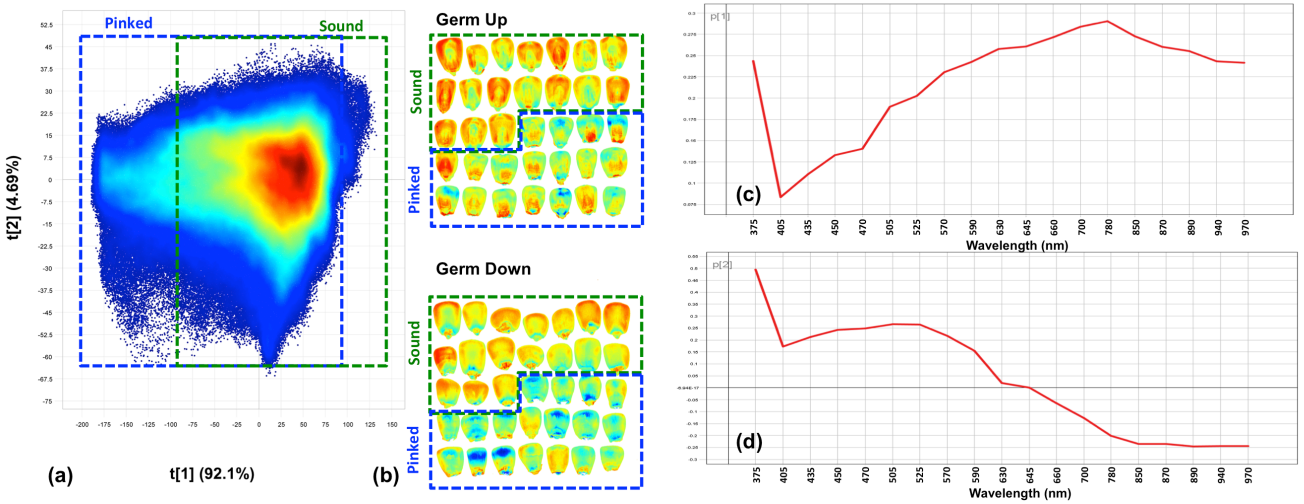

Figure 4.8 Not separated pinked class vs. sound class in (a) the PCA scores plot (PC1 vs. PC2) and (b) the PCA scores image (PC1). Loadings line plots for (c) PC1; and (d) PC2 do not represent the general loadings line trend observed over most of the 13 PCA analyses.

A small number of PCs were needed to explain a sufficient proportion of the variance, with all 13 analyses requiring only up to PC2 or PC3. This was likely due to the relatively small number of spectral variables, namely 19 wavebands. The PCA scores image and scores plot were inspected to provide an overview of the data. The scores are calculated by projecting each sample object according to two or more PCs, and give an indication of the relationship between sample objects (Amigo *et al.*, 2013). In the score plots, similarity between pixel spectra was shown as clusters, while colour shading indicates the density of points (see Fig. 4.4a). The most commonly used scores plot is that of PC1 vs. PC2, as these are the two directions along which the dataset exhibits the largest variances. In the score images, the similarity in samples was displayed using a heat map, where similar colour shading represents similar score values. The sound class was generally associated with positive scores (hot colour shading), while the undesirable material class was associated with negative scores (cool colour shading), as demonstrated clearly in Fig. 4.4b. This occurred for 11 of the 13 analyses, with only two exceptions, namely the sunflower and screenings (Fig. 4.7) classes, in which the inverse occurred. It is important to note that both scores and loadings are vectors, and their projection as either positive or negative is random and non-specific. However, it is important to note that the pixels/objects with positive score values are always associated with high absorbance at spectral variables in positive loading line peaks.

Many analyses produced well-separated classes in the score plots (PC1 vs. PC2) and score images (PC1). These 9 classes included the defective white maize classes *Fusarium* damage, heat damage (Fig. 4.4a and b) and water damage; yellow maize class (Fig. 4.6a and b); and all 5 foreign matter classes, e.g. wheat (Fig 5a and b). In these cases, the colours (positive and negative score values) of the kernels in the two classes were clearly different in the scores image. The scores plot has clusters in areas of the PC space that were clearly associated with each class, and clusters shared little overlap.

Other analyses showed less separation. The analysis of screenings (Fig. 4.7a and b) gave a partial separation with less defined clusters in the score plots. Some colour shading variation was visible in the scores image. The classes *Diplodia* damage (Fig. 4.8a and b), rodent damage, and pinked maize (Fig. 4.9a and b) gave poor separations, where no trend was evident by visual assessment. The differences between the classes were not discernible in the scores image, while, in the scores plot, pixels from both classes' kernels were projected in the same areas with no separate clusters. In Fig. 4.8a, the single cluster was oriented around the origin of the plot, indicating little separation along both PC1 and PC2. Inspecting the noisy score images of PC2 and PC3 (not shown) did not yield any further evidence of class separation.

A range spanning the UV, visible and the lower NIR region was used to characterise the samples. The majority of the loading lines exhibited high positive and negative loadings at variable 1 (UV) and variable 2 (violet), respectively, followed by a smooth positive curve in the visible region (variables 3-13), and ending with the NIR region (variables 14-18) gradually descending into the

negative loadings with a minimum value at variable 19 (refer to Table 4.1 for wavebands). This pattern was common among the following 9 classes: *Fusarium* damage, *Diplodia* damage, water damage, heat damage, rodent damage, plant material, wheat, sorghum and soy. An example of the PC1 loading line plot pattern is given shown in the results of heat damage (Fig. 4.4c) and *Diplodia* damage (Fig. 4.8c). Variable 1 (375 nm) was the most positive loading, while variable 2 (405 nm) was the most negative loading. Absorbance in this region is associated with moieties commonly termed chromophores (Weeranantanaphan and Downey, 2010). Specifically, conjugated C=C bonds are the most likely constituents to absorb in the region of 375-405 nm, particularly in biomaterials such as maize (Mohan, 2004). These are common in many chemical components, such as proteins or unsaturated fats, and thus it is difficult to pinpoint the exact source of variation contributing to these loadings.

The visible region contributes lower weighted loadings, although variables 6 (505 nm), 7 (525 nm), 8 (570 nm) and 9 (590 nm) may be considered important. These wavelengths are associated with green, yellow and orange colour. The heat damage defect may be used as an example to demonstrate these findings. This defect causes dramatic discolouration of the white maize to a dark yellow/orange colour (Fig. 4.1). The sound class objects were generally projected with positive scores, indicating that the sound kernels had higher pseudo-absorbance responses at these visible variables than the heat damaged kernels. The heat damaged kernels did not absorb at these wavelengths, with the reflected wavelengths causing their dark yellow/orange colour. Sound class was associated with positive scores throughout the study, and was thus associated with higher absorbance in these yellow/orange/red visible regions in cases following the typical loadings line pattern. Most other defects also cause yellow (heat damage), brown (*Diplodia* damage; and water damage), and purple (*Fusarium* damage) discolouration. The sound white maize was also visually whiter in colour than all foreign matter classes.

Lastly, negative loadings in the final NIR bands, mostly from variable 17 (890 nm), 18 (940 nm) and 19 (970 nm), are observed where the kernel's chemical constituents absorb. The peaks at 890 and 940 nm are both related to C–H stretch third overtone, and are associated likely to unsaturated fat and saturated fat, respectively (Ropodi *et al.*, 2013; Dissing, 2011; Panagou *et al.*, 2014). The peak at 970 nm is related to the second overtone band of water (Osborne *et al.*, 1993). Some chemical differences between the classes can be deduced by considering the nature of the undesirable material. Screenings are broken white maize kernels, named as they fall through the sieving screen used during manual grading. These fragments are likely to have a very different mean spectrum compared to the sound class, as no fragment contains all kernel components in the correct proportions that combine to form a sound class spectral signature. Similarly, the rodent damaged kernels result from rodents consuming only the nutritious germ from the kernel, leaving the endosperm intact. Thus, the rodent damaged class is mostly starch and protein, while the sound class contains the oil-rich germ, giving rise to clear chemical differences. The differences between

the foreign matter classes are to be expected, as the different commodities all have unique chemical compositions and, thus, spectral signatures. The chemical difference between the white maize defects is much more nuanced, as the defective kernel was once a sound white maize kernel before undergoing defect onset. A large proportion of the original chemical components remain intact, and modelling based on only defect symptoms can be more challenging. The chemical differences arise due to constituents such as protein, starch and sugar, and oil changing concentration. The fungal defects, *Fusarium* and *Diplodia* damage, would exhibit the depletion of constituents. In a study using the NIR region for *Fusarium* infection of maize detection by Williams *et al.* (2012), the main cause of variation was found to be changes in starch and protein composition or contents. The actual presence of fungal growth did not contribute to the variation in the NIR region. Heat damage occurs when kernels are scorched and discoloured during drying, while water damage occurs when the kernels are exposed to wet conditions pre-harvest producing a brown ring on the lower regions of the kernel attached to the cob (Fig. 4.1) (Symons and Shahin, 2008). For the heat and water damage classes, the defect onset or stress triggers physiochemical changes within the kernel, causing chemical component changes, as well as the obvious colour change.

The sunflower and screenings classes' loading line plots displayed the inverse of the typical pattern. This indicates that the scores were projected inversely along PC1, but that the model was based on the same variables with the similar relative weightings. The only deviation was that the peak at variable 1 did not appear for the screenings class. This inverse projection may be observed in the scores plot and scores image of screenings class vs. sound class (Fig. 4.7), where the sound class was uncharacteristically associated with the negative PC1 scores (blue colour). As previously mentioned, the inverse projection of the projection of scores and loadings does not indicate anything specific. As both the scores and loadings were inverted, the sound class remained associated with the loadings from the visible spectral variables.

The PC2 loadings line plots also followed a similar pattern across most classes. The two classes that displayed the inverse pattern for PC1 were not inverted for PC2. The pattern was common among the following 11 classes: *Fusarium* damage, *Diplodia* damage, water damage, heat damage, rodent damage, screenings, plant material, wheat, sorghum, soy and sunflower. Heat damage (Fig. 4.4d), wheat (Fig. 4.5d) and *Diplodia* damage (Fig. 4.8d) PC2 loadings line plots demonstrate the characteristic shape. Variables 1 (375 nm) and 2 (405 nm) exhibited sharp peaks similar to PC1. New variables to aid discrimination along PC2 were variables 4 (450 nm) and 5 (470 nm), associated with blue colour, and variables 13 (700 nm) and 14 (780 nm), upper red and lower NIR variables, appearing in prominent positive and negative peaks, respectively. The screenings class (Fig. 4.7d) did vary slightly from this trend and had a relatively flat PC2 loadings line for variables beyond 1 and 2, thus indicating that the colour and NIR wavelength variables did not contribute largely to the separation in PC2.

The only exceptions to the PC1 and PC2 loadings line pattern were yellow maize (Fig. 4.6c and d) and pinked maize (Fig. 4.9c and d). Yellow maize (Fig. 4.6d) exhibited the same peaks at variable 1 and 2 in PC1 (58.3% SS), however the loadings from the visible and NIR regions deviated. The variables 6 (505 nm) and 7 (525 nm), associated with green and yellow colour, appeared as positive loadings. The yellow maize kernels have a bright yellow colour (Fig. 4.1) due to the presence of carotenoids, such as beta-carotene, and thus weighting these variables to distinguish between the classes was expected (Eckhoff and Paulsen, 2012). There were very low contributions from the remainder of the spectral variables. Unusually, PC2 (27.8% SS) contributed more highly than most analyses' PC2. The prominent variables 8-13 (570-700 nm) span the yellow to red range of the visible region. Again, the significant colour difference between the two classes played a major role in classification, as observed in both PC1 and PC2. The NIR region, particularly variables 18 (940 nm) and 19 (970 nm), contributed largely. White and yellow maize vary considerably in chemical composition (Cauvain and Young, 2009). These differences in chemical constituent composition and content are indicated by the NIR region variables. The combination of colour and chemical information allowed for a well-separated PCA classification.

The other class to not follow the general loading line pattern was pinked white maize. The shape of the PC1 loading line of pinked maize (92.1% SS) (Fig. 4.9c) was different from all the other analyses, as it only exists in the positive loading space. Positive peaks appear at variables 1 (375 nm), 13 (700 nm) and 14 (780 nm). The visible variables are associated with red colour. The pinked maize kernels have a prominent pink/red colour (Fig. 4.1), and it is expected that this region of the visible spectrum would contribute largely. The NIR region (850-970 nm) contributed considerably to PC1 as variables 15-19. This indicates that there were chemical differences identified not only related to colour. The loading line of PC2 (4.69% SS) (Fig. 4.9d) was largely discriminated by the UV (positive loadings) and NIR (negative loadings) regions, variable 1 and variables 15-19, respectively. Despite the high contributions from PC1 and PC2, the separation of the two classes is not achieved using PCA pixel-wise classification (Fig. 4.9a and b). Instead, it seems that kernel components common to both classes are giving rise to the observed differences, rather than differences between the classes. The variables in the NIR region were utilised to differentiate between the kernels' oil-rich germ, proteinaceous hard-endosperm, and starchy soft-endosperm. This demonstrates that a pixel-wise approach applied to heterogeneous samples is clouded by intra-class variation and inter-class commonality. The importance of applying object-wise analysis in this study is highlighted.

Partial Least Squares Discriminant Analysis

The object-wise PLS-DA models were calibrated with 17 sound kernels and 18 undesirable material kernels/objects, using the germ-up and -down side of each kernel. These models were validated by applying the model to a independent validation mosaic image, As shown in Table 4.2, 8 of the 13

analyses achieved 100% classification accuracy, including: heat damage; screenings; yellow maize; plant material; wheat; sorghum; soy; and sunflower. The remaining classes ranked in descending order are as follows: water damage; pinked maize; *Fusarium* damage; rodent damage; and *Diplodia* damage. The classes heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are used to illustrate the PLS-DA results, as given in Figs. 4.10-4.15.

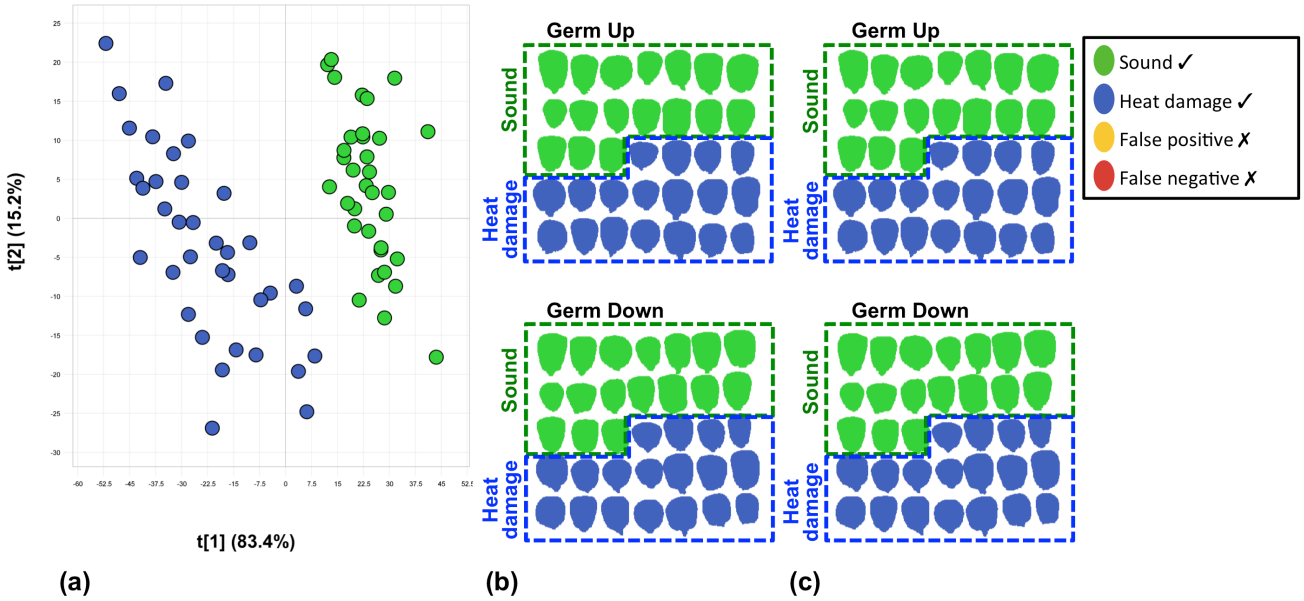

Figure 4.9 Perfectly classified heat damage class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

Figure 4.10 Perfectly classified wheat class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

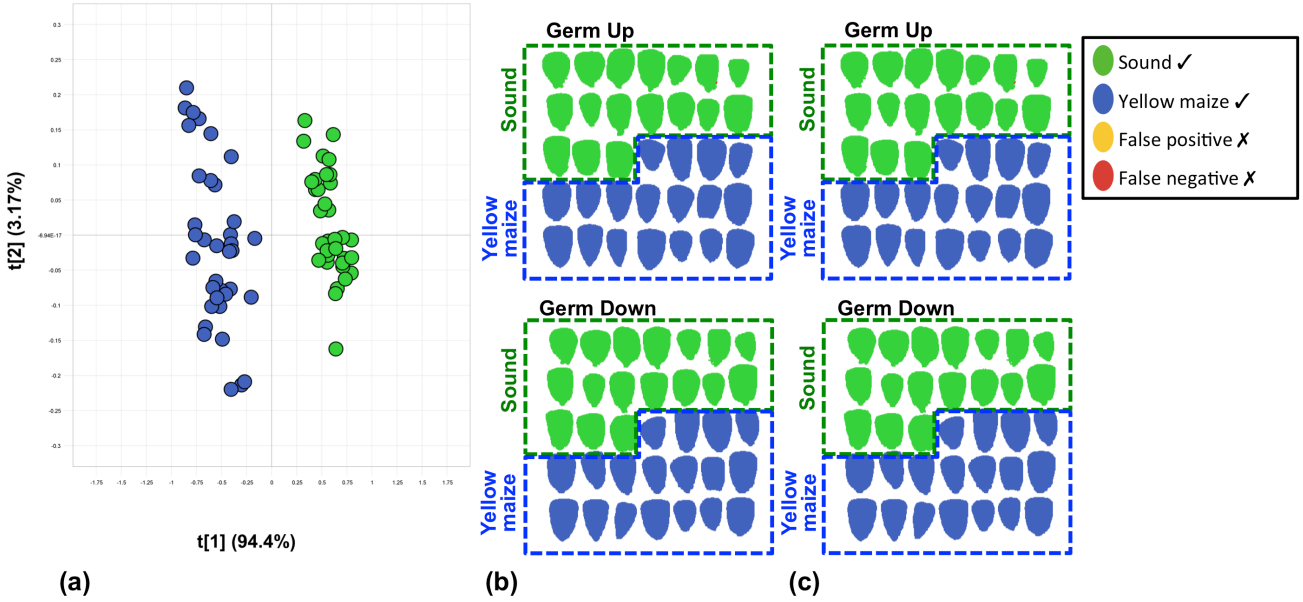

Figure 4.11 Perfectly classified yellow maize class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

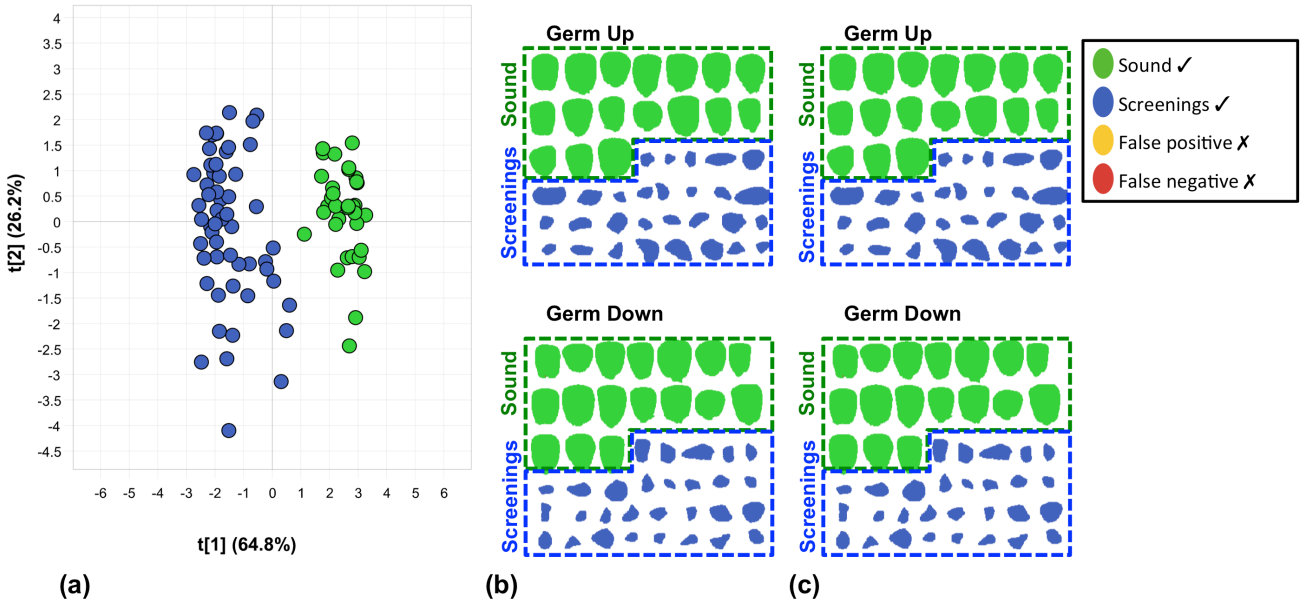

Figure 4.12 Perfectly classified screenings class vs. sound class (100% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

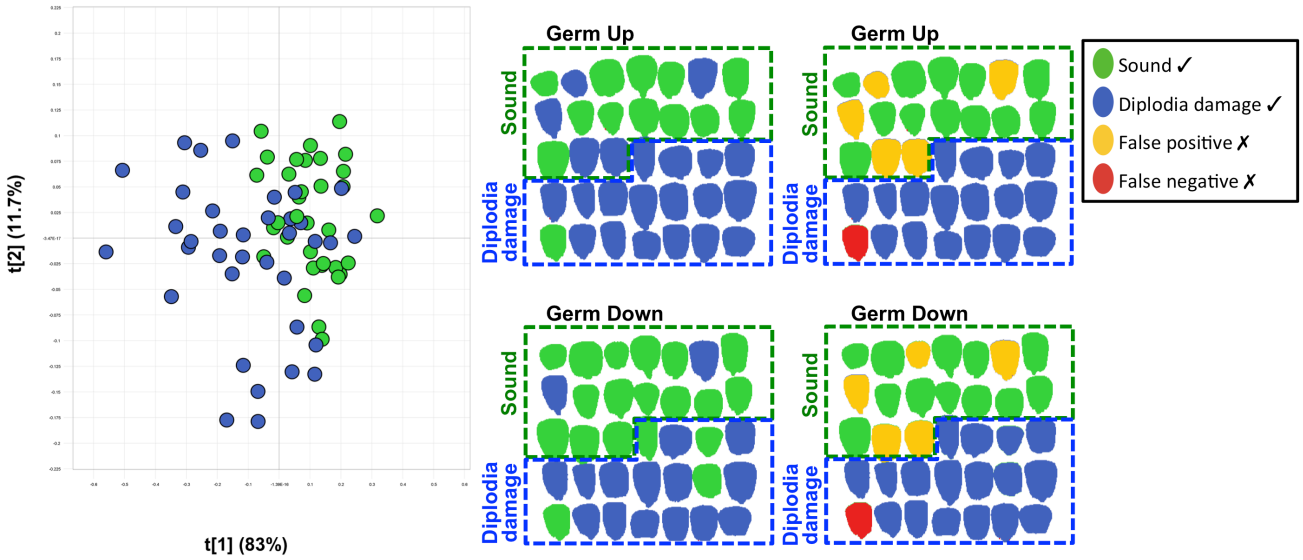

Figure 4.13 Poorly classified *Diplodia* damage class vs. sound class (82.86% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

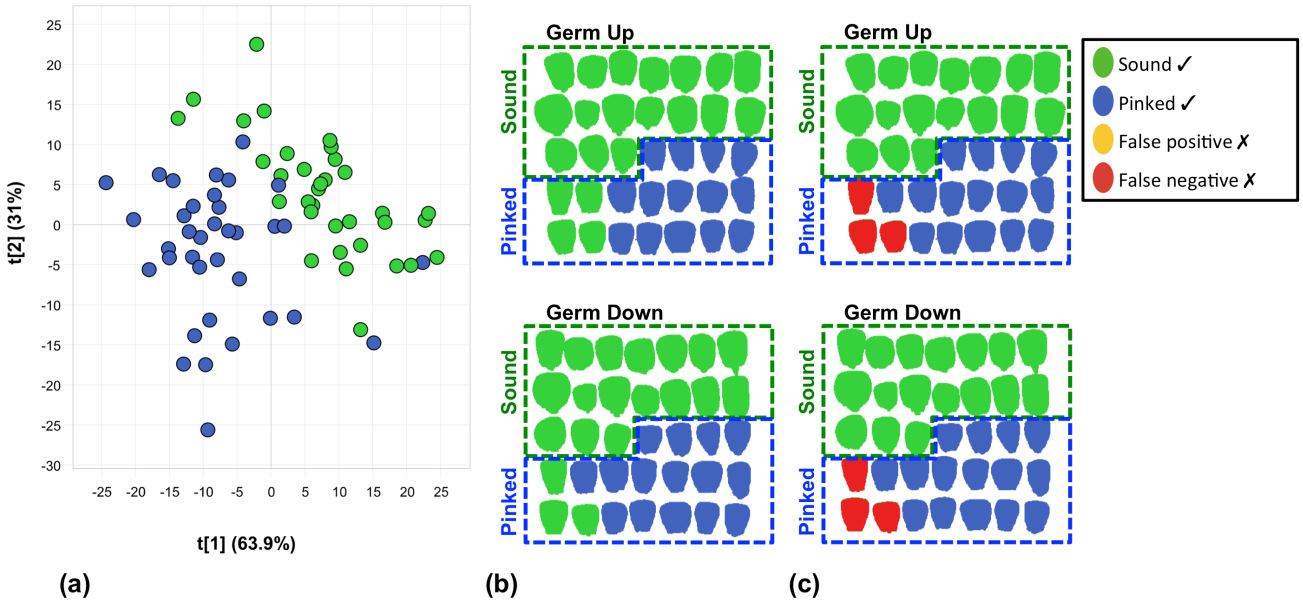

Figure 4.14 Moderately well classified pinked class vs. sound class (91.43% classification accuracy). (a) PLS-DA scores plot (PLS factor 1 vs. 2); (b) Unprocessed classification image; and (c) Overall classification image.

Table 4.2 Results of PLS-DA model calibration and validation for the separation of sound white maize class from 13 undesirable material classes

Undesirable Material	R^2	Q^2	PLS factors	Classification accuracy (%)	False negatives (%)	False positives (%)	Sensitivity (%)	Specificity (%)
Fusarium damage	0.836	0.810	3	88.57	11.43	0.00	77.78	100.00
Diplodia damage	0.459	0.351	3	82.86	2.86	14.29	94.44	70.59
Water damaged	0.912	0.897	2	97.14	0.00	2.86	100.00	94.12
Rodent damaged	0.813	0.766	3	88.57	2.86	8.57	94.44	82.35
Heat damaged	0.891	0.876	3	100.00	0.00	0.00	100.00	100.00
Screenings	0.893	0.886	2	100.00	0.00	0.00	100.00	100.00
Pinked maize	0.790	0.754	3	91.43	8.57	0.00	83.33	100.00
Yellow maize	0.955	0.947	2	100.00	0.00	0.00	100.00	100.00
Plant material	0.951	0.943	2	100.00	0.00	0.00	100.00	100.00
Wheat	0.983	0.982	2	100.00	0.00	0.00	100.00	100.00
Sorghum	0.995	0.994	2	100.00	0.00	0.00	100.00	100.00
Soy	0.972	0.956	2	100.00	0.00	0.00	100.00	100.00
Sunflower	0.986	0.947	2	100.00	0.00	0.00	100.00	100.00

R^2: Coefficient of determination

Q^2: Cross-validated coefficient of determination

PCs: Principal components

An object-wise PLS-DA approach (Kucheryavskiy, 2013) was chosen above the more commonly used pixel-wise approach following the findings in Chapter 3. A substantial proportion the chemical content and visual qualities of a sound white maize kernel and a defective white maize kernel were identical. Both contain the same constituents, distributed similarly, and have comparable colour, shape and dimensions. Thus, pinpointing the subtle differences between the two was challenging at times. Furthermore, the differences between the classes were often clouded by differences within the classes. For instance, grain hardness, cultivar, or severity of defect, which varied substantially within the classes. An object-wise approach was specifically developed for cases where objects from different classes have many similar pixels (Kucheryavskiy, 2013), and was found to be effective in this application. As a kernel with a small area of defect onset must be entirely classified as undesirable material, it is more sensible to use the object-wise approach, and the aim of classifying whole kernels was achieved more directly.

The PLS-DA scores plot shows the projection of the validation objects. The PLS-DA scores plots for heat damage, wheat, yellow maize, screenings, *Diplodia* damage, and pinked classes vs. sound class are given in Fig. 4.10-4.15a. The separations across PLS factor 1 are generally quite distinct, particularly in the foreign matter classes, such as wheat (Fig. 4.11a), and yellow maize (Fig. 4.12a). In these cases, the % SS value for PLS factor 1 was very high (*ca.* 95%). These classes were separated only along PLS factor 1, while only intra-class separation occurred along PLS factor 2 (*ca*. 5%). All foreign matter classes, and the yellow maize class were fully separated in the PLS-DA scores plots.

The defective white maize classes, illustrated using heat damage (Fig. 4.10a), screenings (Fig. 4.13a), *Diplodia* damage (Fig. 4.14a) and pinked maize (Fig. 4.15a), had higher % SS in PLS factor 2 (*ca*. 15%) than the previously discussed classes. This gives an indication that PLS factor 1 cannot achieve clear separation alone. As seen in Fig. 4.10a, the sound class is associated with the positive scores of both PLS factor 1 and 2, while the heat damage class is associated with the negative scores. Clearly separated scores plots were observed for classes water damage, heat damage, and screenings. The *Fusarium* damage class exhibited an overlap by three objects, while separations were not achieved in classes *Diplodia* damage, rodent damage and pinked maize, where the overlapping between classes occurred.

The unaltered classification images display the classification of each object, where green is sound and blue the respective undesirable materials. All objects in the green (sound) box should be coloured green, while all objects in the blue (defect) box should be coloured blue. If objects appear as the incorrect colour, it was determined if a false positive (yellow) or false negative (red) occurred. The unprocessed and overall classification images are given as Fig. 4.10-4.15 b and c, respectively. As expected, the classes with clear separations in the PLS-DA scores plots achieved perfect classifications. All foreign matter classes and the yellow maize class produced perfect classification images, and thus there was no difference between the unprocessed and overall classification

images. Some of the defective white maize classes produced unprocessed classification images with errors. The images thus had to be processed and totalled to produce the overall classification image. The class with the poorest classification was *Diplodia* damage, as can be seen in Fig. 4.14b and c, and exhibited several false positives and false negatives.

Model performance was evaluated using the coefficient of determination (R^2) and cross-validated coefficient of determination (Q^2). The coefficient of determination and cross-validated coefficient of determination describes the fit of the model, where the cross-validated coefficient was calculated based on full cross-validation. The R^2 and Q^2 values were generally quite high, except the *Diplodia* class (0.459 and 0.351, respectively). Few PLS factors were required during PLS-DA model calibration, likely due to the low number of spectral variables. To evaluate the classification result, the classification accuracy, false positive error, false negative error, specificity and sensitivity were calculated. The classification accuracy, false positive error and false negative error demonstrate the classification performance of the model when applied to an independent validation data set. All foreign matter classes, yellow maize and two defective white maize classes (heat damage and screenings) were perfectly classified. All errors occurred during the analysis of the remaining defective white maize classes *Fusarium* damage, *Diplodia* damage, pinked maize, water damage and rodent damage. The model can also be defined using the specificity and sensitivity. The sensitivity and specificity of all of the perfectly classified classes was 100%, while the remaining defective white maize classes yielded lower specificity and sensitivity, due to their classification errors.

CONCLUSION

With the use of multispectral imaging, it was possible to separate sound white maize from the 13 major classes of undesirable materials encountered during industry grading practices. Of the 13 analyses conducted, the exploratory PCA analyses of 9 classes vs. sound class yielded well-separated results. The remaining 4 analyses produced less defined class separations with much overlap between classes in the PCA space. During the PLS-DA modelling, the use of object-wise analysis helped overcome the issue of classes' pixels overlapping, and suited the study's objective to classify a whole kernel as sound or defective. The two-way PLS-DA classification models yielded high coefficients of determination (0.79-0.99) and classification accuracies (83-100%). Perfect classifications were achieved in 8 of the 13 analyses, including all foreign matter classes, yellow maize and two defective white maize classes. As calibration and validation of the models was conducted using independent samples, the results have demonstrated the accuracy of the models. The multispectral method proposed in this study shows advantages to industry as it shows a shorter image acquisition and analysis time, and lower cost than hyperspectral imaging. The non-destructive technique is objective, an advantage over the subjective manual grading methods

currently used. As the method utilised two-way classifications only, the step-wise combination of all models, for instance through hierarchical modelling methods, is recommended for developing a more sophisticated system. The method is highly suitable for the separation of white maize from other cereal commodities and foreign matter, as perfect classification results were achieved for these classes. Furthermore, it shows promise for the separation of sound white maize from the common white maize defects, considering the satisfactory classification results in these classes. However, the results achieved using the hyperspectral imaging instrument (Chapter 3) were generally higher. The grade allocation of maize determines the market price, and even the small errors observed could have a large economic effect in industry. Further improvement of models' classification accuracies and robustness may be achieved using a multispectral instrument with a waveband selection specific to cereal technology applications, as opposed to the pre-built instrument used. These improvements should be investigated ensure the development of an optimal multispectral imaging based automated grading system.

REFERENCES

Amigo, J. M., Martí, I. & Gowen, A. (2013). Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality. In: *Data Handling in Science and Technology*. Pp. 343-370. Amsterdam, The Netherlands: Elsevier Science.

Andresen, M. S., Dissing, B. S. & Løje, H. (2013). Quality assessment of butter cookies applying multispectral imaging. *Food Science & Nutrition,* **1,** 315-323.

Ariana, D., Guyer, D. E. & Shrestha, B. (2006). Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. *Computers and Electronics in Agriculture,* **50,** 148-161.

Berman, M., Connor, P. M., Whitbourn, L. B., Coward, D. A., Osborne, B. G. & Southan, M. D. (2007). Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **15**, 351-358.

Bodevin, S., Larsen, T. G., Lok, F., Carstensen, J. M., Jørgensen, K. & Skadhauge, B. (2009). A rapid non-destructive method for quantification of fungal infection on barley and malt. In: 32nd EBC Congress.

Cauvain, S. P. & Young, L. S. (2009). Maize. In: *The ICC handbook of cereals, flour, dough & product testing: methods and applications*. Pp. 198-206. Lancaster, USA: DEStech Publications, Inc.

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., Bonifazi, G. & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology,* **144,** 64-71.

Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Dissing, B. S. (2011). New vision technology for multidimensional quality monitoring of food processes. Denmark: Technical University of Denmark.

Dissing, B. S., Clemmesen, L. H., Løje, H., Ersbøll, B. K. & Adler-Nissen, J. (2009). Temporal reflectance changes in vegetables. In: *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference*. Pp. 1917-1922. IEEE.

Eckhoff, S. R. & Paulsen, M. R. (2012). Maize. In: *Cereal grain quality* (edited by R. Henry & P. Kettlewell). Pp. 77-112. London, U.K.: Springer Science & Business Media.

Esbensen, K. & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems,* **7,** 67-86.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems,* **120,** 126-135.

Liu, C., Liu, W., Lu, X., Chen, W., Yang, J. & Zheng, L. (2014a). Nondestructive determination of transgenic Bacillus thuringiensis rice seeds (*Oryza sativa L.*) using multispectral imaging and chemometric methods. *Food Chemistry,* **153,** 87-93.

Liu, C., Liu, W., Lu, X., Ma, F., Chen, W., Yang, J. & Zheng, L. (2014b). Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. *PloS one,* **9,** 1-8.

Liu, J., Cao, Y., Wang, Q., Pan, W., Ma, F., Liu, C., Chen, W., Yang, J. & Zheng, L. (2016a). Rapid and non-destructive identification of water-injected beef samples using multispectral imaging analysis. *Food Chemistry,* **190,** 938-943.

Liu, W., Liu, C., Ma, F., Lu, X., Yang, J. & Zheng, L. (2016b). Online Variety Discrimination of Rice Seeds Using Multispectral Imaging and Chemometric Methods. *Journal of Applied Spectroscopy*, 1-7.

Ljungqvist, M. G., Dissing, B. S., Nielsen, M. E., Ersbøll, B. K., Clemmensen, L. H. & Frosch, S. (2012). Classification of astaxanthin colouration of salmonid fish using spectral imaging and tricolour measurement. Technical Report 08, Technical University of Denmark, Informatics.

Mohan, J. (2004). Ultraviolet Spectroscopy. In: *Organic spectroscopy: principles and applications*. Pp. 119-187. Harrow, UK: CRC Press.

Næs, T., Isaksson, T., Fearn, T. & Davies, T. (2002). Data compression by PCR and PLS. In: *A user-friendly guide to multivariate calibration and classification*. Pp. 27-38. Chichester, UK: NIR Publications.

Osborne, B. G., Fearn, T. & Hindle, P. H. (1993). Practical NIR spectroscopy with applications in food and beverage analysis. Longman scientific and technical.

Panagou, E. Z., Papadopoulou, O., Carstensen, J. M. & Nychas, G. J. E. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *International Journal of Food Microbiology,* **174,** 1-11.

Ropodi, A., Panagou, E. Z. & Nychas, G. J. (2013). Assessment of microbiological quality and authenticity of minced meat using multispectral image analysis. In: *8th International Conference on Predictive Modelling in Food, Paris-France, Oral Presentation*.

Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z. & Nychas, G. J. E. (2015). Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International,* **67,** 12-18.

Shrestha, S., Deleuran, L. C., Olesen, M. H. & Gislum, R. (2015). Use of Multispectral Imaging in Varietal Identification of Tomato. *Sensors,* **15,** 4496-4512.

Symons, S. J. & Shahin, M. A. (2008). Quality Evaluation of Corn/Maize. In: *Computer Vision Technology for Food Quality Evaluation* (edited by D. W. Sun). Pp. 401-424. Canada: Elsevier.

Trinderup, C. H., Dahl, A. L., Jensen, K., Carstensen, J. M. & Conradsen, K. (2013). A Comparison of Meat Colour Measurements From a Colorimeter and Multispectral Images. In: *Picking up good vibrations*. Pp. 749-757.

Wang, L., Sun, D. W., Pu, H. & Zhu, Z. (2015). Application of Hyperspectral Imaging to Discriminate the Variety of Maize Seeds. *Food Analytical Methods,* **9,** 1-10.

Weeranantanaphan, J. & Downey, G. (2010). Identity confirmation of a branded, fermented cereal product by UV spectroscopy: A feasibility study involving a Trappist beer. *Journal of the Institute of Brewing,* **116,** 56-61.

Williams, P. J., Geladi, P., Britz, T. J. & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science,* **55,** 272-278.

Williams, P. J. & Kucheryavskiy, S. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry,* **209,** 131-8.

CHAPTER 5
GENERAL DISCUSSION AND CONCLUSION

Maize (*Zea mays L.)* is a cereal crop grown widely in South Africa and in many parts of the world. It is a staple food in many developing countries and is used in many diverse forms for human and animal consumption. Grading is an important step in maize evaluation for determining the condition of the grain and a fair market price. The South Africa grading regulations outline a simple manual process for maize grading (Department of Agriculture, 2009). A grader obtains a 150 g sample of the maize consignment and visually identifies any undesirable materials from a list of categories provided in the legislation. These materials are weighed according to category and compared to an upper limit stipulated for each grade. An issue with the current grading methods is the limited human capacity to conduct consistent manual inspection (Lorente *et al.*, 2012). In complex tasks like cereal grading, it was shown that the inconsistency increases when numerous properties are evaluated simultaneously. This study aimed to investigate a viable alternative to manual maize grading that could provide the maize industry with a reliable automated grading solution in the near future. Near infrared (NIR) hyperspectral imaging has demonstrated potential in various cereal classification studies (Wang *et al.*, 2015; Williams *et al.*, 2012; Bauriegel *et al.*, 2011). The main drawback of the technique is that huge datasets are generated from image acquisition that can impede rapid processing for industry purposes. The cost of the instrument is also substantial. The full potential of hyperspectral imaging was first explored before attempting to achieve similar results on a simple, rapid and less expensive multispectral instrument.

NIR hyperspectral imaging and multispectral imaging techniques were combined with multivariate data analysis to differentiate sound white maize from various undesirable materials. These materials included 13 classes of defective white maize, other cereal commodities and foreign matter. The two imaging techniques were compared for their ability to achieve high classification accuracies and coefficients of determination (R^2). The hyperspectral instrument acquired detailed spectral information in a narrower range than the multispectral instrument and was well suited for detailed laboratory research study. It achieved better classification overall, with object-wise classification accuracies ranging 98-100%, but was not rapid enough and too expensive for current on-line industry implementation. The multispectral instrument acquired spectral measurements in a handful of points spread widely over a broad range. Imaging was simple and quick, ideal for implementation in industrial settings. However, the multispectral system available for this study was pre-built and did not incorporate any of the same wavebands as the hyperspectral instrument. Classifications with consistently high accuracy were not achieved, with classification accuracies ranging 63-100%, lower for most classes than the hyperspectral results. For improved results, it is recommended to investigate building a multispectral instrument according to the prominent

wavelengths identified in the hyperspectral imaging research. As hyperspectral imaging achieved excellent results, it is possible that, by utilising a range including the identified important wavelength bands, a better-suited multispectral instrument could be developed for the specific purpose of maize grading. The hyperspectral instrument operates in the NIR region, and thus the information gathered is related to the chemical composition of the samples. Important spectral features identified for the hyperspectral imaging study were 1219 and 1476 (associated with starch), 1941 (associated with protein), and 2117 (associated with moisture). The multispectral instrument utilised a wide range including the ultraviolet (UV), visible and NIR regions, and thus both visible and chemical differences played a role. With 12 of the 19 wavelengths from the visible region, colour played the principle role in separating classes. The wavebands 505, 525, 570 and 590 nm were identified as the most important visible variables and relate to green, yellow and orange colour. The NIR wavebands 890 and 940 nm (associated with fat), and 970 nm (associated with water) were also identified as important features.

While the study compared the capabilities of hyperspectral and multispectral imaging, the use of different data analysis approaches were explored. Pixel-wise and object-wise approaches were applied to principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). The pixel-wise analysis is the traditional approach to analysing spectral images, however did not suit the study's aim of separating whole kernels. Little research has compared the two approaches. Rather than simply choosing object-wise analysis, both approaches were explored in the first section of the study, the hyperspectral imaging research. The advantage of the richness of pixel-wise analysis was observed in the PCA results. The PCA loadings were seemingly unaffected by the approach used, but the large number of points in the score plots and images of the pixel-wise results allowed better understanding of the data. However, the object-wise approach gave substantially better results during PLS-DA classification. It was concluded that when progressing to the multispectral imaging research, a pixel-wise approach to PCA was best suited for imaging pre-processing and exploration, while object-wise analysis gives superior separation during PLS-DA.

The 13 classes used in the study relate to the South African grading legislation, which outlines 4 main categories of undesirable materials. Either the main category or several sub-categories were used as classes as follows, with the maximum level (% w/w) allowed for WM1, the best white maize grade: Defective white maize (7%) (heat damage, water damage, screenings/broken kernels, *Fusarium* fungal damage and *Diplodia* fungal damage); Pinked white maize (20%); Other colour maize (i.e. yellow maize) (3%); and Foreign materials (0.3%) (wheat, soy, sunflower seeds, sorghum and maize plant material). Due to their similar chemical and physical properties, the defective white maize classes were the most challenging classes to separate from sound white maize. These defects appear in varying degrees and a large portion of the affected kernel often remains intact. This results in a sound kernel and defective kernel having very similar overall chemical composition. The majority of misclassifications for both instruments occurred

during the differentiation of these classes. The pinked white maize class is also technically a defective white maize class, but is assigned its own category, as it is a very minor defect and leniently restricted. The pink colour is very subtle and does not seem to cause major chemical changes within the kernels. Thus, this class was difficult to separate from sound white maize and was misclassified at times by both instruments. The other colour maize class is simply yellow maize, as only white and yellow maize are grown and processed in South Africa. The differences between the two classes were significant enough to achieve good separations with both instruments. Lastly, the foreign materials classes were easily discerned from sound white maize, as the chemical composition and physical colour of the classes were vastly different. All of these classes were classified perfectly using both imaging systems.

The capability of hyperspectral and multispectral imaging for distinguishing sound white maize from all undesirable materials common in South Africa has been demonstrated. This work should be extended to progress from simple two-way classifications to a system that can automatically evaluate all 14 classes (sound maize and 13 undesirable materials) simultaneously. It is suggested that this be achieved through multilevel hierarchical modelling of the 13 two-way PLS-DA models. This type of modelling would entail sequentially classifying all objects using the 13 existing PLS-DA models, starting with the most easily discriminated classes (e.g. the foreign materials) and classifying the most challenging class (e.g. pinked white maize) last. For specific use in the maize industry, a multispectral imaging instrument comprising the key spectral variables identified using the NIR hyperspectral instrument (1219, 1476, 1941 and 2117 nm) may be built. By reducing the original number of spectral factors from 256 to 4, the acquisition time, processing time and cost of the instrument will be minimised. Lastly, the current legislation bases results on weight. A single maize kernel weighs approximately 0.25 g and occupied 250-400 pixels using the hyperspectral imaging instrument. The automated system would need to be calibrated to transfer the number of pixels in each object to an estimated weight. This would allow the calculation and assignment of white maize grade WM1, WM2 or WM3.

In conclusion, spectral imaging techniques give a rapid, reliable and accurate description of maize's internal and external quality and safety status. It can evaluate information both visible and hidden to the naked eye. These techniques have the potential to offer the maize industry an alternative to the current manual grading processes, and ensure a more consistent determination of grain condition and fair market price.

REFERENCES

Department of Agriculture (2009). Regulations relating to the grading, packing and marking of maize intended for sale in the Republic of South Africa. In: *Agricultural Product Standards Act (Act No. 119 of 1990).*

Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U. & Herppich, W. B. (2011). Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture,* **75,** 304-312.

Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O. L. & Blasco, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology,* **5,** 1121-1142.

Wang, L., Sun, D. W., Pu, H. & Zhu, Z. (2015). Application of Hyperspectral Imaging to Discriminate the Variety of Maize Seeds. *Food Analytical Methods,* **9,** 1-10.

Williams, P. J., Geladi, P., Britz, T. J. & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science,* **55,** 272-278.

ADDENDUM 1

SOUTH AFRICAN MAIZE GRADING REGULATIONS

Regulations relating to the grading, packing and marking of maize products intended for sale in the Republic of South Africa

DEPARTMENT OF AGRICULTURE

AGRICULTURAL PRODUCT STANDARDS ACT, 1990
(ACT No. 119 OF 1990)

REGULATIONS RELATING TO THE GRADING, PACKING AND MARKING OF MAIZE INTENDED FOR SALE IN THE REPUBLIC OF SOUTH AFRICA

The Minister of Agriculture, acting under section 15 of the Agricultural Product Standards Act, 1990 (Act No. 119 of 1990), made the regulations in the Schedule; and

(a) determined that the said regulations shall come into operation on date of publication.

SCHEDULE

Definitions

1. In these regulations any word or expression to which a meaning has been assigned in the Act shall have that meaning and, unless the context otherwise indicates -

"**bag**" means a bag manufactured from -

(a) jute or phormium or a mixture of jute and phormium; or

(b) polypropylene that complies with SABS specification CKS632;

"**bulk container**" means any vehicle or container in which bulk maize is stored or transported;

"**consignment**" means -

(a) a quantity of maize of the same class, which belongs to the same owner, delivered at any one time under cover of the same consignment note, delivery note or receipt note, or delivered by the same vehicle or bulk container, or loaded from the same bin of a grain elevator or from a ship's hold; or

(b) in the case where a quantity referred to in paragraph (a), is subdivided into different grades, each such quantity of each of the different grades;

"**container**" means a bag or a bulk container;

"**defective maize kernels**" means maize kernels and pieces of maize kernels that;

(a) are shrivelled, obviously immature, frost-damaged, mouldy or chalky;

(b) discoloured by external factors such as water and sun: Provided that discoloration on both sides of the maize kernel limited to less than a quarter from the bottom tip of the maize kernel shall not be considered as defective;

(c) have sprouted, including kernels of which the shoot (plumule) in the germ is visibly discoloured;

(d) have cavities in the germ or endosperm caused by insects or rodents;

(e) are visibly soiled (smeared) or contaminated by smut, fire, soil, smoke or coal-dust;

(f) can pass through the 6,35 mm round-hole sieve;

(g) are of subspecies other than *Zea mays indentata* or *Zea mays indurata*.

Provided that -

(i) irregularity of shape and size of maize kernels shall not affect the grading thereof; and

(ii) chipped or cracked maize kernels or pieces of maize kernels which are in a sound condition and which appear in a sample of maize, but which do not pass through a 6,35 mm round-hole sieve, shall not be regarded as defective maize kernels under these regulations.

"**discoloured maize kernels**" means maize kernels that are as a result of environmental conditions more than 50% discoloured on both sides of the kernel, excluding pinked maize kernels;

"**foreign matter**" means all matter other than maize, glass, stone above the sieve , coal, dung or metal;

"**frost damaged**" means maize kernels that are covered with wrinkles on both sides of the kernel to the crown and have a pearl-like appearance. Maize kernels of which the bran is flaking is considered frost damaged if signs of frost damage are present;

"**heat damaged**" means kernels that are as a result of external heat or internal fermentation affected with excess moisture and have at least one of the following characteristics:

(a) Kernels or pieces of kernels that are completely brown, dark-brown or amber discoloured.

(b) Kernels of which the germ has dark-brown to black discoloration.

"**insect**" in relation to maize, means any live insect which is injurious to stored grain, irrespective of the stage of development of the insect;

"**maize**" means the threshed kernels or pieces of kernels of the plants of *Zea mays indurata* and *Zea mays indentata* or one or more crossings of the two types;

"**mouldy**" means kernels or pieces of kernels that;

(a) are visibly infected by fungi and are characterised by black, blue, green, yellow or white fungi growth anywhere on the kernel, or are characterised by fungi growth underneath the bran layer of the kernel;

(b) are infected by ear-rot and are characterised by red, pink or brown discolorations. The kernels are partially to completely infected;

"**poisonous seeds**" means seeds or part of seeds of plant species that may in terms of the Foodstuffs, Cosmetics and Disinfectants Act, 1972 (Act No. 54 of 1972) represent a hazard to human or animal health when consumed, including seeds of *Argemone mexicana, Convolvulus* spp., *Crotalaria* spp., *Datura* spp., *Ipomoea* spp. *Lolium temulentum*, *Ricinus communis* or *Xanthium* spp.;

"**other colour maize kernels**" in relation to -

(a) white maize, means maize kernels or pieces of maize kernels of which the endosperm as a result of genetic (characteristics) composition have another colour than white, excluding pinked maize kernels;

(b) yellow maize, means maize kernels or pieces of maize kernels of which the endosperm as a result of genetic (characteristics) composition have another colour than yellow ;

"**pinked maize kernels**" means kernels and pieces of kernels of white maize of which the pericarp or part

thereof is shaded red or pink in colour;

"**shrivelled or obviously immature maize kernels**" means maize kernels with a thin and shrunken appearance;

"**sprouted maize kernels**" means maize kernels which have sprouted so far that developing roots and/or sprouts are clearly visible, or the shoot (plumule) in the germ is visibly discoloured;

"**the Act**" means the Agricultural Product Standards Act, 1990 (Act No. 119 of 1990); and

"**the 6,35 mm round-hole sieve**" means a sieve ;

(a) with a flat metal sheet bottom of 1,0 mm thickness perforated with round holes of 6,35 mm in diameter that are arranged with the centres of the holes at the points of intersection of an equilateral triangular grid with a pitch of 8 mm;

(b) of which the upper surface of the bottom is smooth;

(c) the frame of which is at least 40 mm high;

(d) with the inner width of at least 200 mm and the inner length of at least 300 mm, or, in the case of a circular sieve, the inner diameter of at least 278 mm; and

(e) that fits onto a tray with a solid bottom and must be at least 20 mm above the bottom of the tray.

Restrictions on sale of maize

2. (1) No person shall sell maize in the Republic of South Africa -

(a) unless the maize is sold according to the classes set out in regulation 3;

(b) unless the maize complies with the standards for the class concerned set out in regulation 4;

(c) unless the maize complies with the grades of maize and the standards for grades, where applicable, set out in regulations 5 and 6 respectively;

(d) unless the maize is packed in accordance with the packing requirements set out in regulation 7;

(e) unless the containers or sale documents, as the case may be, are marked in accordance with the marking requirements set out in regulation 8; and

(f) if such maize contains a substance that renders it unfit for human consumption or for processing into or utilisation thereof as food or feed.

(2) The Executive Officer may grant written exemption, entirely or partially to any person on such conditions as he or she may deem necessary, from the provisions of subregulation 1:Provided that such exemption is done in terms of section 3 (1) (c) of the Act.

QUALITY STANDARDS

Classes of maize

3. The classes of Maize shall be -

(a) Class White Maize;

(b) Class Yellow Maize; and

(c) Class Other Maize.

Standards for classes of maize

4. (1) A consignment of maize shall be classified as Class White Maize if -

(a) subject to the allowable deviation in respect of other colour maize kernels that apply to the different grades of white maize, it consists of maize the endosperm of which is by nature white in colour; and

(b) it complies with the standards for one of the grades of white maize set out in regulation 6.

(2) A consignment of maize shall be classified as Class Yellow Maize if -

(a) subject to the allowable deviation in respect of other colour maize kernels that apply to the different grades of yellow maize, it consists of maize the endosperm of which is by nature yellow in colour; and

(b) it complies with the standards for one of the grades of yellow maize set out in regulation 6.

(3) A consignment of maize shall be classified as Class Other Maize if the consignment does not comply with the standards for Class White Maize or Class Yellow Maize.

Grades of maize

5. (1) Maize of the Class White Maize shall be graded as WM1, WM2 or WM3.

(2) Maize of the Class Yellow Maize shall be graded as YM1, YM2 or YM3.

(3) No grades are determined for Class Other Maize.

Standards for grades of Class White Maize and Class Yellow Maize

6. All grades of maize -

(a) shall be free from a musty, sour or other undesired odour;

(b) shall be free from glass, metal, coal or dung;

(c) shall be free from a substance which renders it unfit for human consumption or for processing into or utilisation thereof as food or feed;

(d) shall be free from insects;

(e) shall be free from stones which cannot pass through the 6,35 mm round-hole sieve;

(f) shall contain not more than one gram of stones, which can pass through the 6,35 mm round-hole sieve, per 10 kg;

(g) shall contain not more poisonous seeds than permitted in terms of the Foodstuffs, Cosmetics and Disinfectants Act, 1972 (Act No. 54 of 1972);

(h) shall have a moisture content of not more than 14 per cent; and

(i) shall not exceed the maximum percentage of permissible deviation as determined in the table in the Annexure for each grade.

II
PACKING AND MARKING REQUIREMENTS

Packing requirements

7. Maize of different classes and grades shall be packed in different containers.

Marking requirements

8. Each container or the accompanying sales documents of a consignment of maize shall be marked or endorsed with -

(a) the class of the maize; and

(b) the grade, in the case of Class White Maize or Class Yellow Maize.

III
SAMPLING

Obtaining sample

9. (1) A sample of a consignment of maize shall -

(a) in the case of maize delivered in bags and subject to regulation 10, be obtained by sampling at least ten per cent of the bags, chosen from that consignment at random, with a bag probe: Provided that at least 25 bags in a consignment shall be sampled and where a consignment consists of less than 25 bags, all the bags in that consignment shall be sampled; and

(b) in the case of maize delivered in bulk and subject to regulation 10, be obtained by sampling that consignment throughout the whole depth of the layer, in at least six different places, chosen at random in that bulk quantity, with a bulk sampling apparatus.

(2) The collective sample obtained in subregulation (1) (a) or (b) shall -

(a) have a total mass of at least 10 kg; and

(b) be thoroughly mixed before further dividing.

(3) If it is suspected that the sample referred to in subregulation (1)(a) is not representative of that consignment, an additional five per cent of the remaining bags, chosen from that consignment at random, shall be emptied into a suitable bulk container and sampled in the manner contemplated in subregulation (1)(b).

(4) A sample taken in terms of these regulations shall be deemed representative of the consignment from which it was taken.

Sampling if contents differ

10. (1) If, after an examination of the maize taken from different bags in a consignment in terms of regulation 9(1), it appears that the contents of those bags differ substantially -

(a) the bags concerned shall be placed separately;

(b) all the bags in the consignment concerned shall be sampled in order to do such separation; and

(c) each group of bags with similar contents in that consignment shall for the purposes of these regulations be deemed to be a separate consignment.

(2) If, after the discharge of a consignment of maize in bulk has commenced, it is suspected that the consignment could be of a class or grade other than that determined by means of the initial sampling, the discharge shall immediately be stopped and the part of the consignment remaining in the bulk container, as well as the grain that is already in the collecting tray, shall be sampled anew with a bulk sampling apparatus or by catching at least 20 samples at regular intervals throughout the whole offloading period with a suitable container from the stream of grain that is flowing in bulk.

Working sample

11. A working sample shall be obtained by dividing the representative sample of the consignment according to the ICC 101/1 (Approved 1982) method.

IV
DETERMINATION OF OTHER SUBSTANCES

Determination of undesirable odours and harmful substances

12. A sample of a consignment of maize shall be sensorial assessed or chemically analysed in order to determine -

(a) whether it has a musty, sour or other undesirable odour; and

(b) whether it contains a substance that renders the maize unfit for human consumption or for processing into or for utilisation as food or feed.

Determination of glass, metal, coal, dung, stone, poisonous seed and insect content

13. A consignment of maize shall be sensorial assessed and a sample of that consignment shall be sensorial assessed and sorted by hand in order to determine whether the sample contains glass, metal, coal, dung, insects, stones and poisonous seeds.

Determination of percentage of foreign matter

14. The percentage of foreign matter in a consignment of maize shall be determined as follows:

(a) Obtain a working sample with a mass of at least 150g from the sample of the consignment.

(b) Remove all foreign matter from the working sample and determine the mass thereof.

(c) Express the mass thus determined as a percentage of the total mass of the working sample.

(d) Such percentage shall represent the percentage of foreign matter in the consignment concerned.

V
MAIZE KERNELS

Determination of percentage of defective maize kernels

15. The percentage of defective maize kernels in a consignment of maize shall be determined as follows:

(a) Obtain a working sample with a mass of at least 150g from the sample of the consignment.

(b) Place the working sample on the 6, 35 mm round-hole sieve that is fitted onto a matching tray, sieve it in such a manner that all the material on the sieve passes at least 20 times over the entire surface of the sieve and determine the mass of the maize kernels and pieces of maize kernels that passed through the sieve.

(c) Express the mass determined in terms of paragraph (b) as a percentage of the total mass of the working sample.

(d) Remove all defective maize kernels from that part of the working sample remaining on the sieve and determine the mass of the defective maize kernels thus removed.

(e) Express the mass determined in terms of paragraph (d) as a percentage of the total mass of the working sample.

(f) Calculate the sum total of the masses determined in terms of paragraphs (b) and (d).

(g) Express the combined mass calculated in terms of paragraph (f) as a percentage of the total mass of the working sample.

(h) In the case of yellow maize the percentage obtained -

(i) in terms of paragraph (c), represents the percentage of defective maize kernels in the consignment concerned, which can pass through the 6,35 mm round-hole sieve; and

(ii) in terms of paragraph (e), represents the percentage of defective maize kernels in the consignment concerned, which can not pass through the 6,35 mm round-hole sieve.

(i) In the case of white maize, the percentage obtained in terms of paragraph (g) represents the percentage of defective maize kernels in the consignment concerned.

Determination of percentage of other colour maize kernels

16. The percentage of other colour maize kernels in a consignment of maize shall be determined as follows:

(a) Obtain a working sample with a mass of at least 150g from the sample of the consignment.

(b) Remove all other colour maize kernels from the working sample and determine the mass thereof.

(c) Express the mass thus determined as a percentage of the total mass of the working sample.

(d) Such percentage shall represent the percentage of other colour maize kernels in the consignment concerned.

Determination of percentage of pinked maize kernels

17. The percentage of pinked maize kernels in a consignment of maize shall be determined as follows:

(a) Obtain a working sample with a mass of at least 150g from the sample of the consignment.

(b) Remove all pinked maize kernels from the working sample and determine the mass thereof.

(c) Express the mass thus determined as a percentage of the total mass of the working sample.

(d) Such percentage shall represent the percentage of pinked maize kernels in the consignment concerned.

VI
MOISTURE CONTENT

Determination of moisture content

18. The moisture content of a consignment of maize may be determined according to any suitable method: Provided that the results thus obtained are in accordance with the maximum permissible deviation for a class 1 moisture meter as detailed in ISO 7700/1-1994(E) based on the results of the 72 hour, 103°C oven dried method (AACC Method 44/15A/1981).

VII
OFFENCE AND PENALTIES

19. Any person who contravenes or fails to comply with any provision of these regulations shall be guilty of an offence and upon conviction be liable to a fine or imprisonment in terms of section 11 of the Act.

ANNEXURE/AANHANGSEL
TABLE/TABEL

STANDARDS FOR GRADES OF CLASS WHITE MAIZE AND CLASS YELLOW MAIZE/
STANDAARDE VIR GRADE VAN KLAS WITMIELIES EN KLAS GEELMIELIES

	Maximum permissible deviation/ Maksimum toelaatbare afwyking					
Deviation/Afwyking	White maize/ Witmielies			Yellow maize/ Geelmielies		
	WM1	WM2	WM3	YM1	YM2	YM3
1	2	3	4	5	6	7
1. Foreign matter [regulation 14] Vreemde voorwerpe [regulasie 14]	0,3%	0,5%	0,75%	0,3%	0,5%	0,75%
2. Defective maize kernels, above and below the 6,35 mm round-hole sieve [regulations 15,]/ Gebrekkige mieliepitte, bo en onder die 6,35 mm-rondegatsif [regulasies 15,]	7%	13%	30%	9%	20%	30%
3. Other colour maiz kernels [regulation 16]/Mieliepitte van 'n ander kleur [regulasie 16]	3%	6%	10%	3%	6%	10%
4. Deviations referred to in items 1, 2, and 3 collectively: Provided that the deviations are individually within the specified limits/Afwykings in items 1, 2, en 3 bedoel, gesamentlik: Met dien verstande dat die afwykings individueel binne die gespesifiseerde perke is	8%	16%	30%	9%	20%	30%
5. Pinked maize kernels [regulation 17]/ Verrooide mieliepitte [regulasie 17]	20%	20%	*	*	*	*

* Not specified/Nie gespesifiseer nie.